THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR Mary Cronin

TITLE OF THESIS Construct Validity and the Assessment of Reading Processes

DEGREE FOR WHICH THESIS WAS PRESENTED PhD

YEAR THIS DEGREE GRANTED 1982

THE UNIVERSITY OF ALBERTA

CONSTRUCT VALIDITY AND THE ASSESSMENT OF READING PROCESSES

by

©

MARY C. CRONIN

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF ELEMENTARY EDUCATION

EDMONTON, ALBERTA

FALL, 1982

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled Construct Validity and the Assessment of Reading Processes

submitted by Mary Cronin

in partial fulfilment of the requirements for the degree of Doctor of Philosphy

in Elementary Education

ABSTRACT

This is a study of the assessment of reading processes where particular emphasis is placed on establishing the construct validity of the underlying theory of reading. Following a brief survey of the history of testing (measurement) in reading, a theory consisting of eight cognitive and linguistic processes employed by the reader while constructing or reconstructing the author's message is delineated. This theory is developed as an explication of the reading assessment situation specifying two methods of assessing reading processes, viz., miscue and recall analyses. A model of test construction incorporating concerns of construct validity was then devised. Both the theory of reading processes and the model of test construction were utilized to construct the instrument -- the Assessment of Reading Processes or ARP --designed to collect validational evidence. To this end, a factor analytical study was carried out to gather evidence of reading processes from a sample of 102 grade-four readers. Because this was an exploratory study, no definite hypotheses were deduced from the theory ; however, six dependent variables were included in the analysis to act as outside criteria for the processes. The outcome of the factor analysis consisting of eight oblique (Promax) factors was interpreted and related to similar research findings. Then the implications of the factors for reading and identifying specific processes were presented. Finally, suggestions for further research and educational practice were made. All in all, it was concluded that a certain measure of construct validity was established for the theory of the assessment of reading processes and the instrument designed to measure those processes, the ARP.

ACKNOWLEDGEMENTS

Among the persons at the University of Alberta who made this work possible, I wish to first thank William Fagan, my supervisor, for his sure guidance, friendship, and unflinching support ; also Jean Robertson and Len Stewin, who read an earlier draft, for the many insightful revisions they provided. Sincere thanks are also due to Grace Malicky, Regna Darnell, and Patricia McFetridge for their interest and stimulating questions. I am grateful to Tom Maguire, Division of Educational Research, who was always available to offer help with the statistical problems.

I am indebted to the Edmonton Catholic School System who permitted me to do research in the schools ; to their administrative staff who provided me with normative data ; to the school principals and teachers ; and last but not least, to the children who read the stories.

Finally, a special word of thanks is due to Laurie Walker, from Mount Saint Vincent University, Halifax, who as my external examiner provided useful last-minute comments and revisions.

TABLE OF CONTENTS

Chapter Page

Chapter | Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION TO THE STUDY

Overview

> Reading is a very elaborate procedure, involving a weighing of many elements in a sentence, their organization in the proper relations one to another, the selection of certain of their connotations and the rejection of others, and the cooperation of many forces to determine final response (Thorndike, 1917, p. 323).

In 1917 Thorndike, one of the founding fathers of achievement testing, expressed this "elaborate" view of the intricate processes involved in reading comprehension at the sentence level in his famous paper, "Reading as reasoning: A study of mistakes in paragraph reading". But his concept of paragraph comprehension is not limited to understanding individual sentences; it also encompasses the abstraction and organization of the meaning of larger units of discourse:

> The mind is assailed as it were by every word in the paragraph. It must select, repress, soften, emphasize, correlate and organize (p. 329).

Implicit in these quotations is the conceptualization of reading as interaction between the reader's mind and the text as the reader selects, organizes and rejects aspects of printed information.

It is remarkable that almost half a century should have elapsed before a similar notion of the complexity of the mental processes involved in comprehending print would direct the main body of research in reading. In the intervening years reading was generally conceived of as a set of hierarhical sub-skills. And as a consequence of the push for universal literacy which

gathered increased momentum after the First World War, emphasis was focused on teaching reading rather than understanding how a child learned to read. Furthermore, these trends resulted in a dichotomized concept of reading as the ability to identify words and the ability to comprehend (see Smith, 1965 for the historical background to this period). The skills approach to reading often resulted in models of reading that focused on the teaching rather than the learning and explanation of reading (e.g., Gray, 1960) :

> The skills approach to an explanation of reading comprehension suffers form a basic confusion over what domains of behavior and cognitive activity can rightfully be characterized as reading comprehension. This confusion has led to the inclusion of reading comprehension skills that tend to be global and vague and that fail to distinguish between those processes that are specific to reading and those that are very general cognitive processes (Simons, 1970, p. 5).

However, since the 1960's there has been a genuine effort to describe what the reader does when he reads: how he selects cues from print, integrates them with his stored knowledge in order to construct the author's message while simultaneously committing that newly created semantic structure to memory. Yet, regardless of these significant theoretical advances few efforts have been made to measure both the cognitive and linguistic processes involved in reading. Efforts at measuring reading processes have been carried out for research purposes and are, therefore, restricted in their applicability for more general use. At present, there is no comprehensive readily available instrument to measure the quantity or quality of processing the reader engages in. Now, as in the days

of Thorndike, such a measure depends on the existence of an integrated theory of reading comprehension. In several respects, educational measurement has not significantly changed since the pioneering days of Thorndike and his students at Columbia University (see Chapter 2 for a detailed account of the history of testing): "The part that shows as change is the one-ninth of the iceberg that is above the water level, and the part that remains much the same is the vast underwater bulk" (Thorndike, 1971, p. 3).

One of the most persistent criticisms of traditional standardized tests of reading is that they are not theoretically based (e.g., Farr, 1969, 1970; Simons, 1970 and Walmsley, 1975); Kingston (1970) complains that "it is the rare (test) author who justifies his selection of materials on theoretical grounds" (p. 230). Thus, a fundamental problem arises with all norm-referenced reading tests, both word-identification and comprehension; it is not at all clear what they are measuring. For example, they may be measuring in addition to, or in place of, reading ability:

1. a student's memory skill on tests where the student is not allowed to look back to the passage (Carroll, 1972; Davis, 1962), or
2. a student's familiarity with the passage content resulting in the ability to answer comprehension questions without reading the passage (Tuinman, 1971).

The majority of reading tests are constructed from a descriptive analysis of reading as a set of sub-skills, or of a domain of content or objectives; therefore, the type of behavior classified as

reading is not clearly defined.

Norm-referenced tests are effective only for the purposes for which they were designed, that is, to differentiate among pupils' reading but definitely not to measure an individual's knowledge or to characterize how he processes print. To use these tests for purposes other than differentiation or to attempt to make instructional decisions based on a standardized score is to seriously misinterpret and misuse these tests (see Farr, 1970). A score on a norm-referenced test allows one to make statements about a pupil's performance relative to that of the norming population. Nothing can be inferred from a standardized score, much less a sub-test score (Farr, 1968), about how a student reads; in other words, such a score makes statements about relative reading product, not reading process. Yet, despite the fact that norm-referenced tests are useless for diagnostic or absolute measurement purposes,

> in the area of reading instruction, norm-referenced measures are widely used for measuring students' reading status and growth, diagnosing reading difficulties, selecting students for remedial reading programs, evaluating the effectiveness of these programs and so on (Walmsley, 1975, p. 2).

Nonetheless, it must be emphasized that these tests do effectively differentiate and categorize students broadly, e.g., below average/ above average.

Another variety of measures, criterion-referenced or domain-referenced tests, has appeared in the educational measurement arena since the mid-1960's (Berk, 1980; Hambleton et al., 1978; Hively, 1974; and Popham, 1976, 1978, 1980). These two titles refer

to basically the same characteristics of a measure (see Hively, 1974, pp. 140-141), that is, measurement referenced to a specific domain and are often used interchangeably. Even when this problem of duality of title is clarified, confusion still reigns because domains are defined in different ways by each author or test constructor, for example: domains of objectives, tasks or cognitive skills (see Nitko, 1980, for a comprehensive classification).

These tests are designed to yield absolute rather than comparative (with respect to other examinees) measures; the notion of criterion enters in as the intention is to measure performance with reference to a clearly defined area of instruction, competence, etc. Popham's (1978) definition is representative of what many writers feel is the main advantage of criterion- over norm-referenced tests: "A criterion-referenced test is used to ascertain an individual's status with respect to a well defined behavioral domain" (p. 95). Furthermore, for the majority of writers, whether they attack the problem deductively (Baker, 1974) or inductively (Hively, 1974), specifying a domain is almost synonymous with giving a "lucid description of the nature of the items on the test" (Popham, 1980, p. 16). It is this "lucid description" that allows the investigator to interpret the signifigance of an examinee's test score.

The emergence of criterion-referenced tests awakened hopes that finally there was a technically feasible means of measuring theoretically meaningful units of reading behavior. This concern is conveyed by this excerpt from Farr (1969):

> The major obstacle in testing and measurement today is a lack of clear understanding of what the reading process entails. Until a theoretical construct of reading is developed and substantiated, the value of testing devices will remain extremely limited. However, once reading is defined, the avenues for test development will broaden: it will be possible to develop criterion-referenced tests geared to assess how well an individual reads on the basis of what reading is rather than on the basis of how others perform (p. 215).

However, these early hopes have not been fulfilled. Technical issues in referencing test items to clearly specified and finite domains still delay more rapid progress (Hambleton et al., 1978). In an effort to overcome the major problem of specifying the limits of complex domains, reading is fragmented into smaller areas of content and skills (e.g., Walmsley, 1975, 1978). But, even more fundamentally, theoreticians of these tests are more concerned with specifying domains than with elucidating process accounts of the interaction between the examinee and the test item (see Chapter 4 for a more detailed argument).

For the remainder of this study the terms criterion-referenced and domain-referenced will be utilized interchangeably.

Norm-referenced tests and criterion-referenced tests refer to the general interpretation of the test score rather than to techniques for measuring comprehension. Carroll (1972) classified the various methods used in the literature to measure comprehension of both oral and written language. In a subsequent discussion, Trabasso remarked that hardly any of these methods had been devised based on an explicit model of comprehension (Carroll, 1972, p. 25). Among these atheoretical techniques were recall and

question answering, which are two of the methods of assessment employed in <u>informal reading inventories</u> that are commonly used to diagnose strengths and weaknesses and to help the classroom teacher and reading specialist make instructional decisions. Even though the format of informal inventories makes them suitable for assessing reading processes, in general they are not used for this purpose. The teacher observes the child as he reads, noticing his various oral reading errors; the amount of passage content comprehended is then ascertained by questioning or an unaided recall. Yet, without a theoretical explanation it is impossible to translate the various error or comprehension scores into statements about process. However, one of the main reasons why this method for measuring comprehension could be modified to measure process is that "reading is widely conceived. The interest is not on mere pronunciation of words, but also in the manipulation of ideas which are represented by these words" (Johnson and Kress, 1965, p. 15).

Lack of a theoretical basis, however, is probably at the root of the ad hoc system of item selection and setting of reading levels. The passages are not systematically chosen either for content or for structure; but are generally chosen because they resemble the content of a basal reading series. Informal reading inventories commonly have a norm-referenced aspect as they establish an independent, instructional and frustration level of reading for the student. The criteria for setting these three levels have been severely criticized by Powell (1969), who complains that the levels are arbitrarily set too high.

When discussing methods of measuring comprehension, Carroll (1972) did not mention a qualitative analysis of oral reading errors, generally known as miscue analysis. This omission, however, is understandable as it was in 1972 that a systematic account of this system of analysis was published (Goodman and Burke, 1972). This publication, in fact, marked an initial step in the measurement of reading process; although few processes were named, a relatively simple method for inferring such processes was born. Psycholinguistic theory formed the theoretical structure for building the instrument and drawing general processing inferences. Writing in 1969, Farr anticipated that psycholinguistics might open the door to measuring how the reader processed print:

> Perhaps the psycholinguistics approach will provide a more viable definition of reading and lead to a more solid base for test construction. It may well be that research will find, as the proponents of psycholinguistic theory have suggested, that attempts to define reading sub-skills on a group basis are fruitless. In that case, measurement in reading would have to be based on whether a reader has a strategy for decoding written messages and whether he understands reading as a communication process rather than whether he can decode written symbols, supply the meaning of words in isolation, or answer multiple-choice questions based on a literal understanding of a selection. Until research is carried out to develop tests which take into account the elements psycholinguistic theorists are finding central to reading ability the teacher will still need to use present sub-tests of reading to evaluate reading ability, but this use of sub-tests must be done cautiously (p. 9).

At present, psycholinguistic theory is quite well developed and has accumulated a considerable body of validating research (see Beebe, 1980, and Haupt and Goldsmith, 1982). As a consequence of its

theoretical stance, however, cue selection or psycholinguistic theory has explained passage comprehension as the processing of a successive chain of sentences.

To complement the psycholinguists, another group of theorists loosely dubbed the "discourse processing theorists" (Beebe, 1981) has emerged even more recently (e.g., Frederiksen, 1977; Kintsch and Van Dijk, 1978; and Mandler and Johnson, 1977). In general this latter group of theorists are concerned with passage structure, its information content and with the reader's a priori knowledge of that structure and content. Their research has shown that recall of passages mirrors both the passage structure and the a priori perceptual framework of the reader (Kintsch and Van Dijk, 1978; and Steffensen et al., 1979). Both the psycholinguistic and discourse processing theories account for reading as the processing of textual information; furthermore both attempt to specify discernable and measureable units of comprehension. These are complementary rather than competing accounts of reading which focus on different aspects of passage comprehension (see the conclusions of Beebe, 1981).

Nevertheless, test constructors such as Anderson et al. (1978) and Walmsley (1975, 1978) complain of the problems involved in constructing a theoretically based test of reading because of lack of an all-encompassing theory of reading. To evade this problem these latter authors have constructed criterion-referenced tests of small subdomains of reading. It is, however, the contention of this author that the general outline of such a theory is available

but lacks integration and cohesion. Once such a synthesis is attempted the path to the construction of tests of meaningful units of reading is close. Furthermore, since the basic building blocks of this proposed theory are process accounts of both word identification and comprehension, the way should then be clear for the construction of a test to measure the reading processes, provided the aid of psychometricians is sought.

A very general, but nevertheless, basic definition of measurement is "rules for assigning numbers to objects in such a way as to represent quantities of attributes" (Nunnally, 1978, p. 3). The "attributes", in this case, cognitive and linguistic processes, are the concern of reading theorists; whereas the characteristics of the "numbers" lie in the domain of the psychometrician. In contrast, the "rules" and how they are formulated cross both areas of expertise. Therefore, to devise a precise and valid measure of reading process it is necessary to bring together two disparate areas of knowledge: (1) knowledge of reading theory, and (2) knowledge of measurement theory-- the latter taking its identity from applied statistics. The lack of integration of these two areas of knowledge in many previous efforts at constructing measuring instruments in reading contributed to their weaknesses. When referring to the flaws of many of the criterion-referenced measures, Walmsley (1975) states:

> One reason for this may lie in the fact that two somewhat separate groups are involved in the development of tests--researchers from the field of statistics and educational measurement and researchers from the field

> of reading theory and instruction. While these two groups may be aware of the developments in their own fields, are they always conversant with developments in each other's fields (p. 7)?

Even theoretically based tests will fail to achieve their purpose if inappropriate statistical techniques are used.

The Measurement of Process

In the past two decades, many behavioral sciences, e.g., anthropology, biology, history and psychology, have undergone a major paradigm shift (Kuhn, 1962/1970) from the study of structures and their taxonomies to the exploration of process or dynamic structures. Even death has been reconceptualized: in holistic medicine it is no longer considered an event diametrically opposed to life, but rather a process which forms a continuum with life (Moore et al., 1980, p. 180ff). Recently, interest in processual accounts of cognition - and its many sub-areas such as memory, learning, or reading - has been expressed by many researchers and theorists (see Clark and Clark, 1977; Dodd and White, 1980; and Swenson, 1978). Traditional theories of intelligence were typically structuralist (e.g., Cattell, 1971; and Guilford, 1968), whereas more recent efforts to conceptualize it focus on a series of processes (Feuerstein, 1979); similar trends are evident in psychobiology where some researchers view the brain as a process (Restak, 1979, p. 20). Belth (1977), a philosopher, describes thinking as the process of "deliberately" constructing a model, analogy, or metaphor "so that we might entertain in some

systematic way events of our experience" (p. xvii).

Perhaps the appeal of process theories is their explanatory power and capacity for elucidating "relationships between otherwise distinct and independent clusters of phenomena" (Estes, 1979, p. 47). Since the aim of science is to explain natural phenomena, furthermore since these phenomena are incessantly changing and interacting, a dynamic process account is more satisfactory for explaining the nature of these constant interactive transformations. In other words, a process account is more efficient for understanding the nature of change or how change may be effected. And since effecting a change in learning is the aim of educators, their arguments in favor of measures of reading process capable of tracking a student's progress, diagnosing strengths and weaknesses, and then making the necessary decisions about instructions, are hardly suprising (e.g., Farr, 1969, 1970; Walmsley, 1975; and Wardrop et al., 1978).

Furthermore, reading processes, such as selecting graphic cues or encoding the abstract semantic information in memory, must operate in equilibrium; faulty processing strategies, at any point in the cognitive operations involved in bringing print to meaning, will affect the overall construction of meaning. Thus, a holistic depiction of how a child processes print is prior to effective instruction. A process method for the dynamic assessment of intelligence has been developed by Feuerstein (1979) with the aim of modifying defective cognitive processing. Intelligence

is thus viewed as a dynamic process rather than a relatively immutable substance: "intellectual functioning is seen to be the expression of a complex interaction of biogenetic, cultural, experiential and emotional factors, and thus reversability of poor intellectual performance may definitely be anticipated" (Feuerstein, 1979, p. 50). Feuerstein is aware not only of the complex nature of the forces determining intelligence but also of the interaction of the cognitive processes themselves. His definition of cognitive functions is similar to that given above for reading process: "To a large extent cognitive function represents autoplasticity - the way in which an organism changes itself in response to the disruption of its equilibrium by sudden changes in the internal or external enviroment" (p. 6).

The Learning Potential Assessment Device (LPAD) is the instrument created to measure the dynamics of intelligence, it attempts to measure the organism's ability to adapt by drawing up an inventory of the examinee's adaptive behavior. To achieve this end, changes are required not only in the test structure and orientation (from a product to a process orientation), but in addition in the examination situation and in interpretation of the results. The testing situation becomes a cooperative adventure between examiner and examinee, where the child is first taught how to carry out the required task but is also given further assistance when necessary. In conventional intelligence tests, average performance is scored and interpreted, whereas in the LPAD great weight is attached to peaks in performance.

The LPAD theory and device are intimately linked to an ongoing program of research in cognitive modifiablity; furthermore, the LPAD presents a paradigm for formulating research questions (Feuerstein, 1979, p. 275, ff). Answers to these questions will bring refinements to the assessment techniques and theory, greater direction for the prescriptive training program to develop effective cognitive processing and ultimately a contribution to science in increased knowledge about the nature of intellectual functioning and cognitive modifiability. Each refinement, direction, or contribution will also help to establish construct validity for the LPAD.

Just as this assessment of intellectual processes is firmly rooted in a corresponding theory of cognitive function, so must a device suited to the measurement of reading processes be grounded in a conceptual explanation of reading as active reader-directed processing. But such a theory is necessary, more specifically, for the following reasons:

(1) Reading processes are covert and cannot be measured directly, therefore they must be inferred from their products or effects. "The inference about a property from its effects involves either an assumption about a relation between effect and property, or a demonstration about the relation between effect and property" (Lorge, 1967, p. 45). Either the assumption or the demonstration is necessarily based on some level of theoretical explanation.

(2) In order to measure any property or trait, some concept or notion of it must first exist. Although reading is a highly complex, covert activity, it is possible to measure it - provided the various reading processes are adequately and accurately mapped out. As Kaplan (1964) responds in answer to the question, "Are there things which are intrinsically immeasureable?":

> For my part I would answer these questions with an unequivocal 'No'...I would say that whether we can measure something depends, not on that thing, but on how we have conceptualized it, on our knowledge of it,...

But Kaplan goes on to state that measurement depends

> above all on the skill and ingenuity which we can bring to bear on the process of measurement which our inquiry can put to use (p. 176).

In the case of the measurement of reading processes, a great deal of "skill and ingenuity" must be invested while devising an effective instrument for inferring process from product. Such an instrument must be based on reading theory, but designed according to techniques developed by psychometricians.

Kaplan's (1964) statement implies that if something exists it can be measured, depending on the scientist's ability to interpret the sensory perception or on the ability of instruments to "approximate more closely the property under observation" (Lorge, 1967, p. 47). However, the converse is not implied; that is, if something is measured it exists. Hence, reading processes could conceivably be measured with little or no basis in reality.

The explication of the links between theory and measuring instrument, and between instrument and the world of reading process, is the function of the construct validation of measures. "To determine construct validity, a measure must fit a theory about the construct; but to use this as evidence it is necessary to assume the theory is true" (Nunnally, 1978, p. 104; emphasis in the text). A paradoxical relationship, therefore, exists between theory and instrument of validation, and measurement. And as a logical consequence, if any measure of construct validity is established for either a theory of reading process or the corresponding measure of that construct of reading, a measure of construct validity is established for both.

A construct "may be relatively simple and narrow, or it may be relatively complex and comprehensive" (Lorge, 1967, p 46). For instance, the construct of reading processes could be intended to refer to each of the reading processes individually, so that each would have its own theoretical sub-structure, e.g., selecting graphic cues, predicting the author's message, or synthesizing ideas abstracted from across the passage, each one supported by distinct theories or at least some form of conceptual networks. Or the construct could be complex and broad where the entire process of gaining meaning from print would be explained as the dynamic interaction of all the individual processes. Such a construct of reading process is adopted in this work.

Problem

McDonald's (1970) following statement could be said to paraphrase the major criticisms of measures of reading: "Measurement is useless without a clear theoretical sub-structure. Unless we know what we are trying to find, the utility of our measures will only be a function of our uncertainty" (p. 10). Therefore, construct validity cannot be established for reading tests without a theory of reading. Various accounts of reading as cognitive and linguistic processing exist at the present but await integration (e.g., Goodman, 1976; Kintsch and Van Dijk, 1978; and Smith, 1978). Once such a synthesis is achieved, work on a comprehensive instrument to measure reading processes may begin.

Yet, if such a theory is to be used effectively to assess reading processes, a theory of the assessment of reading processes needs to be developed. This contextualization of the reading processes within an assessment situation is required because of the dual nature of tests as samples and signs of behavior.

> Because tests are samples, what we know about behavior in general applies to test behavior. Because tests are signs, what we learn from tests can help structure and interpret knowledge of other, more amorphous behavior.
>
> The problem then becomes one of finding items which are significant for practical and theoretical purposes (Loevinger, 1967, p. 85).

It is the purpose of the theory of assessment to outline how the signs of processing may be inferred from reading product, but also

to map out how the samples of reading behavior observed in an assessment situation differ from that behavior in general. In other words, how generalizable are the characteristics of reading processing in the assessment situation?

A measurement instrument may then be constructed to the specification of the theory of assessment of reading processes, utilizing psychometric techniques. But if this instrument is to be valid either for applied or theoretical purposes, some estimates of construct validity must be established. Furthermore, because of the intimate relationship between a scientific theory and the instrument employed to measure it, both are validated and modified simultaneously.

Purpose

Arising from the research problem outlined above, four separate but intertwined goals have been set for this study:

1. To synthesize a theory of reading process and to develop it within an assessment framework.
2. To develop a test theory suited to the construction of a measure of reading processes and the construct validation of both theory and instrument.
3. To construct a method of assessment, utilizing both the theory of reading assessment and the test theory.
4. To construct a factor analytical study in order to collect some validational evidence on the nature of the reading processes.

Focus of the Study

This project in theory building and measurement will focus on the nature of the interaction between reader and text. Therefore, the major emphasis in the theory of reading assessment will be on elucidating the nature of this exchange. Nevertheless, it is accepted that there are many factors, both internal (emotional state) and external (ethnicity of the examiner), which may affect the quality and extent of processing or meaning construction. And it follows also that the focal point of test construction will be on the method of analysis designed to infer process from product, rather than on the content of the assessment items per se; the system of analysis could be used to assess reading processes which occur when using a variety of texts.

Plan of the Study

The remainder of this study will consist of seven chapters. The next chapter will contain a brief outline of the history of measurement in reading. In the third chapter the theory of assessment will be mapped out; and the fourth chapter will consist of a similar exposé of the theory of test construction. The methods employed to construct and revise the measurement items will be presented in chapter five, followed by a description of the factorial validity study in the sixth chapter. The findings of this study will be reported and discussed in the seventh chapter, whereas the final chapter will consist of a summary of the entire

study, and a discussion of the broader theoretical and practical implications.

CHAPTER II

HISTORY OF MEASUREMENT AND TESTING IN READING

Introduction

Consistent attempts have been made to measure the processes and products of reading since the end of the last century. But the testing of reading achievement as we know it today did not exist before the second decade of this century. For the purpose of delineating the general trends in measurement and testing in reading, this portion of time--roughly from about 1870 to the present-- will be divided into four broad historical periods. Similar accounts of the history of psychological testing are often divided into two periods: from roughly 1869, the publication of Sir Francis Galton's early work, Heredity and Genius, to the founding of the Psychometric Society and its journal Psychometrika in the mid-thirties; and the second, from the end of the first period to the present (e.g., Carroll, 1978; and Dubois, 1970).

Before 1910

During this period, the measurement of reading closely approximated trends and innovations in psychological testing (see Jenkinson, 1957, pp. 40-50); in more recent times, however, significant independent advances have been made in the assessment of reading, e.g., miscue analysis. Interest in the study of individual differences in mental ability blossomed and was nurtured

and guided by an intellectual and social context which included the formulation of the theory of evolution, an elitist social ideology later call "Social Darwinism" (Hofstadter, 1944), the end of the Industrial Revolution, and mass immigration to the States (Blum, 1978; and Carroll, 1978). In England Galton attempted to establish simple physical or perceptual measures of "innate" intellectual abilities. Similar trends soon followed in the United States. Cattell founded the Psychological Laboratory at Columbia University in 1891, where the tests "were largely of sensory and motor functions, with related measures of perception, association and memory beginning to appear" (Dubois, 1970, p. 22).

Most of the measures taken in the pioneering days of psychological measurement were in physical units; this was obviously in keeping with the experimental psychology of the day, which emphasized the investigation of the physiological causes of psychological events, e.g., sensation, reaction time, and discrimination. Investigations in reading also tried to measure physical phenomena such as eye-movements. Simple perceptual studies were also carried out which measured the perception of words and letter strings, and "the extent of reading matter perceived during a reading pause" (see Huey, 1908, pp. 15-101). The measurement of reading was then clearly a part of mainstream psychological measurement.

In accordance with the above observations, it is inter-

esting to note that in his retiring address as President of the American Association for the Advancement of Science in 1925, Cattell said that the publication of the results of his research in visual perception in reading in 1885 was a bench mark in psychology as (1) these were the first measurements of individual differences in psychology, and (2) this was the first time the term individual difference was ever used (Poffenberger, 1947, p. 384).

1910 - 1939

About 1910 the notion of measuring educational products swept across the fields of American education and psychology. Teacher's College, Columbia University under the direction of Edward Lee Thorndike, was at the epicentre of the educational testing movement. Many of the new measures of achievement were published by Thorndike's students, among whom were W.S. Gray and A.I. Gates (Dubois, 1970). The influence of Thorndike was, therefore, evident in much of the early work both in the testing and in the study of reading. In 1914, he published his Alpha Test which measured the child's ability to understand word meanings. Furthermore, he went on to publish several articles delineating the problems involved in constructing a reading test (Jenkinson, 1957).

Scores of tests appeared from this time onwards and because of the growing awareness of measurement criteria, care was taken in such attempts to establish their validity and

reliability through stringent standardization procedures. So tests of reading achievement came on the market, many of them complete with "grade norms, measures of reliablity and other necessary information" (Gates, 1921, p. 303). To illustrate this growth, Gates (1921) cites a (then) recent bibliography as containing the titles of eighteen tests of silent reading, three of oral reading, and six for vocabulary (e.g., Gray Oral Reading Test, Monroe Silent Reading Test and Thorndike's Scale for Understanding Sentences). In this review article of reading tests, Gates (1921) also carried out the first of a long list of critiques and analyses of tests (e.g., Farr, 1969; Gates, Bond and Russell, 1939; McCullough, 1958; and Wardrop et al., 1978). It is obvious from the previous chapter of this thesis that his criticism of contemporary tests still holds true today, "...even the more carefully constructed of our educational tests are insufficiently refined for exact individual examination in a short period of time" (p. 281). In addition, Gates (1921) also noted the problem of the distinctiveness of the various sub-skills measured by separate subtests. Again this problem was to plague the theory and measurement of reading until recently. However, underlying theoretical issues were not the main concern at this stage; tests strove to be objective, reliable and norm-referenced. Objectivity and efficiency in scoring were a primary concern of the enitre field of psychological measurement at this phase (see Dubois, 1970).

The concern with objectivity probably ensured that the majority of reading tests would conform to a multiple choice format; in addition this facilitated the development of group measures of reading proficiency. Nevertheless, some tests tried to be all inclusive and diagnostic, and measured as many facets of reading ability as possible.

As Jenkinson (1957) states, "the Holmes Reading Test for Grades 2-12, in 1915, measured the ability to answer factual questions on a story read within a certain time and yielded a comprehension and rate of reading score. Pupils were then asked to reproduce the story in writing" (pp. 42-43), and were further graded on their comprehension and penmanship.

In the 1930's interest shifted from the measurement of general reading ability to the measurement of early predictors of successful reading; in other words, to the testing of reading readiness in grade one. Much of the research in this period was trying to establish measures of those competencies that would guarantee success in reading instruction classes. Similar to measurement in other area of reading, these studies in reading readiness have no theoretical base for choosing successful predictors. The measures or predictors, such as visual discrimination, were usually chosen because they correlated with reading achievement later in grade one; and, again in keeping with the main trend of the period, these tests were norm-referenced.

1940 - 1968

The 1940's heralded two new events in the history of reading measurement and testing: studies of factor analysis carried out on the sub-tests of norm-referenced test, and the construction of a new type of assessment, namely the informal reading inventory.

Factor Analysis. The advent of factor analysis meant that there was now a statistical tool to measure the distinctiveness of the diverse sub-skills of reading. At this time as many as several hundred such skills could be identified by counting the names of sub-tests of different tests. In some cases existing tests were administered to large numbers of students (e.g., Holmes, 1953, 1970; Singer, 1970; and Traxler, 1941); whereas in others special tests were constructed for this purpose (e.g., Davis, 1944). Correlation matrices were obtained from the results, and the principal factors accounting for the variances were factored out and given names. The results obtained from these numerous studies is at best inconclusive (see Lennon, 1970 for a review of literature). The various comprehension skills tested by standardized reading tests were shown to be highly correlated and in general show that a common factor accounted for reading comprehenion. It should be noted that all these studies were carried out with junior high and high school students. However, McCullough (1957) achieved similar results with the administration of tests to grades one, two and four.

There was no theory of reading or empirical evidence to show that the sub-skills measured by these sub-tests did in fact exist; hence these tests never succeeded in gaining construct validity. Theorizing based on data obtained through factor analytical procedures worked itself into a blind alley and contributed directly to the impoverishment of reading theory, because these atheoretical sub-tests were factor-analysed in order to discover a small number of factors or constructs, and these "constructs" were then to be used as the foundations of a theory of reading and for the construction of future tests. Singer's (1970) theory is a prime example of this cat-and-mouse game which existed between reading theory and testing from the 1920's until the mid-60's. During this period reading testing was considered one of the main keys to reading theorizing. As Jenkinson (1969) states, "One of the major techniques used in studying (reading) comprehension has been through the means of tests, and consequently, much of the research in reading has been concerned with the construction and evaluation of reading tests" (p. 45). In recent years this approach to theory building was abandoned.

Informal Reading Inventories (IRI). In an article in the early 1940'2 Betts made the first mention of an IRI (quoted in Walter, 1974). He pointed to the advantages while indicating how the classroom teacher could construct his/her own tests for diagnostic purposes. An IRI consists of a graded set of passages accompanied by a set of comprehension questions.

As the child reads the passage orally, a detailed listing of his errors are made for diagnostic purposes.

Betts published one of the first commercially available informal reading inventories and Smith (1959) published a set of graded passages for inclusion in IRIs. Betts (1954) also set the criteria for establishing the three levels of reading competence obtained from an IRI: independent, instructional, and frustration -- which criteria he obtained from Killgallon's (1942) doctoral dissertation. And so the accepted criteria were that a child had to reach niney-five percent accuracy on word identification to attain his instructional level. "This is the level at which the child should be and can profitably be instructed" (Johnson and Kress, 1965, p. 7). According to Walter (1974), this appears to have been an arbitrary decision; however, he adds, "It should not be forgotten that these criteria have been fairly well established through use by many of the most prominent authorities in the field of reading" (p. 5). Therefore, the conventions used in many IRIs have gained credibility from continuous use since the 1940's.

In the 1960's and 1970's many other IRI's were published. McCracken's (1966) inventory is typical of many of these. He constructed The Standard Reading Inventory using the content of three basal reading series, and applied readability formulas to ascertain the difficulty levels of the passages. McCracken's (1964) study seems to be the most thoroughly conducted investigation of an IRI to date. He field-tested his inventory on a sample of 600 pupils from grades one to six. He reports that his tests placed children within an acceptable

instructional range and notes that standardized tests often place pupils too high, overestimating the instructional level and placing them at frustration level.

Informal inventories are not designed to differentiate between pupils. Instead they attempt to appraise an individual's performance on a particular task without reference to what others do. Their primary purpose, however, is to find the correct level to commence reading instruction. This categorizing into reading levels automatically gives IRIs some of the characteristics of normed tests.

Until the present, informal reading inventories have proven to be useful diagnostic tools, but Walter (1974) lists two reasons to explain why they are not used more widely: (1) they are time-consuming to administer, and (2) many teachers need training in order to administer them effectively.

1969 - 1982

During this period traditional forms of measuring and testing reading have come under severe criticism (e.g., Farr, 1969, 1970; Gibson and Levin, 1975; Simon, 1970; and Wardrop et al., 1978). The work of Kintsch and Vipond (1977) on readability has pointed to the inadequacy of readability measures used on the informal reading inventories. With one noteworthy exception, there have been few advances in the testing of reading. The significant step forward made in methods of assessment--miscue analysis--is rooted firmly in a psycholinguistic theory of reading, and the results of its application have in turn advanced

that theory.

> In the late 1960's and early 1970's Kenneth and Yetta Goodman attempted to synthesize the ideas of psychologists and linguists into a psycholinguistic theory of reading based upon their observations and evaluations of the errors that children made while reading aloud. The scheme devised by the Goodmans allowed investigators, through the qualitative analysis of oral reading errors, to observe the strategies readers were using (Beebe, 1981, pp. 5-6).

Miscue analysis rests on the assumptions that (1) the quality of the errors is more important than the quantity of errors, and (2) the errors or miscues are not just a random response, but represent the reader's systematic attempt to gain meaning from a text. The <u>Reading Miscue Inventory</u> which is the instrument devised to analyse the miscues in terms of graphophonic, syntactic, and semantic similarities with the text word, was published in 1972 (Goodman and Burke, 1972). This enabled the teacher, clinician, or researcher to investigate a student's reading strategies as he reads any text, including informal reading inventories. Thus, with the use of this method of analysis the emphasis is on the measurement of reading process, not product.

Recently, considerable pressure has been brought to bear on the reading field to develop sensitive measures of reading. As a result many other advances may be at hand. Walmsley (1975) attributes this changing state of affairs to three phenomena:

1. There has been a change in educational philosophy. Education is now seen to serve a developmental rather than a selective function.

2. At present, cognitive development is viewed as the aim of education; and level of thought processes, not the difficulty or correctness of the product, as the focus of measurement. This is a view of educational measurement that increasingly rejects norm-referenced tests.
3. Recent political, social, and educational demands for "accountability", "individualized instruction", and "diagnostic and prescriptive" teaching have high-lighted the inadequacies of the presently available reading tests.

The exigencies listed above have given rise to a flurry of activity in the area of criterion-referenced tests (see Berk, 1980; Popham, 1976 and 1978). But concerns have centered only on the demarcation of a domain of content to be measured; and not on establishing a consistent "inference system"* for interpreting the scores (Wardrop et al., 1978), or on the construct validity.

Interest in the measurement of reading has not waned over the years. Now more than ever, there exists among both the educational profession and general public "an extreme desire and interest in finding out how well students are reading" (Farr, 1970, p. 2).

---

* These points will be elaborated in chapter 4: in particular, advances in the conceptualization of validity in psychometric theory will be discussed.

Conclusion

The history of measurement in reading, terminating at the present state of the art, testifies to the serious lag which exists between the theory of reading and how reading is generally measured. The present study is then an effort to redress this balance. As an initial step in this direction, a theory of reading assessment will be outlined in the following chapter.

CHAPTER III

A THEORY OF THE ASSESSMENT OF READING PROCESSES

Introduction

In the first chapter the value of process theories in the understanding of cognitive functioning was discussed. Dodd and White (1980) explain the underpinnings of such theories as follows:

> Process accounts describe the way in which input is treated from the moment it arrives in the sensory system until it enters consciousness and/or affects behavior. Recent process accounts detail how the processing system is organized or structured and how goals affect the way in which input is processed as well as the nature of the processes themselves (p. 8).

This depiction aptly summarizes the process account of reading which will be explicated in the present research. In order to establish construct validity for a process theory of reading it is necessary to gather empirical data. Before these data can be gathered, however, it is also necessary to develop a theory of reading assessment. As a consequence of the theoretical basis it will be possible (1) to abstract rules for the generation of the structure and content of the assessment items; (2) to specify the components of the theoretical domain which may affect the difficulty of the items for readers; (3) to go beyond the given data -- in this case, infer process from product; and (4) to infer aspects or variables within reader processing which contribute to particular characteristics or patterns of performance.

This theory of assessment will also delimit the behavior that can be categorized as reading. Comprehension will act as the outside criterion; therefore reading is viewed as the comprehension of written thoughts that take place when the reader interacts with the print.

With these purposes in mind the remainder of this chapter will be devoted to the explication of a theory of the assessment of reading processes.

A Theory of the Assessment of Reading Processes

Reading takes place whenever the reader constructs the meaning of the text, thus liberating its "potential for meaning" (Fagan, 1978)--in essence, the creation of reality through inference and imagination (Bronowski, 1973). The reader interacts with the text, simultaneously abstracting selected features (e.g., letters, words, or structures), integrating these with his anticipations and world knowledge, allowing the text a certain autonomy but always building a model of the message to fit the specifications of his pyschic state. The encounter between reader and text may take place in an infinite number of different situations. Typically, the situational context defines not only the emotional state of the reader but also his purpose for reading, which then influences processing. In the reading assessment situation this processing is further affected by the presence of the examiner. Therefore, any theory which attempts to explain reading assessment must begin to unravel

the nature of the intricate interaction between reader and text in addition to the influence of the situational variables.

The basic components of the proposed theory of reading assessment are modeled in Figure 1. Reader-text interaction occurs in oral and silent modes of reading in the context of the reader-examiner encounter and the broader situational variables. Three forms of reading product are also specified: oral reading and recalls after oral and silent reading. Although the exchange between reader and test is usually covert, it cannot be assessed except through some variety of product.

Assessment Situation

The purpose of any reading assessment situation is to measure how well a reader reads. Traditionally, assessment has focused on the "what" of reading--assessing proficiency by establishing a level of word recognition or text difficulty for each individual. In contrast, this study does not focus primarily on level of achievement, but rather on how text is processed as the reader constructs or reconstructs the author's meaning. How an investigation chooses to gain insights into the processing of text information to a large extent determines the mode of reading and the method of analysis.

Mode of Reading. It is possible to engage in two different modes of reading--oral and silent--while getting meaning from print. Although for most puroposes, silent reading is the most useful and functional method of reading, extensive

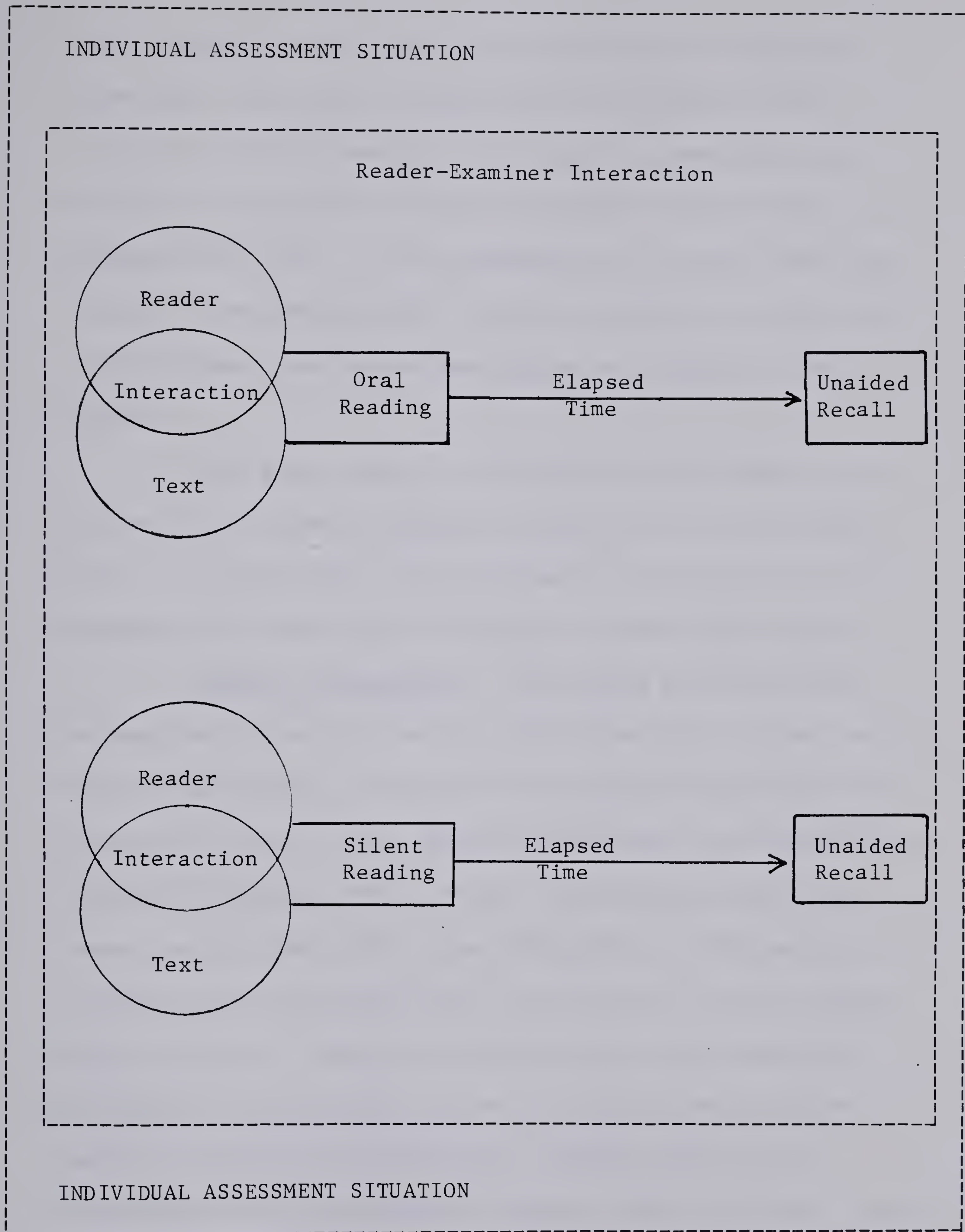

Figure 1. The Reading Assessment Situation

use is made of the oral mode in instructional and assessment situations. According to Kenneth and Yetta Goodman (1977) "an analysis of oral reading offers unique opportunities for the study of linguistic and pyscholinguistic processes and phenomena" (p. 317). If the assessment is to benefit from the insights on processing gained through an analysis of oral reading performance, the reader must engage in a sample of oral reading.

This mode, however, is neither the most usual nor the most efficient means of gaining meaning from the printed page. Hence, it follows that if the assessment of processes is to be complete, the reader must also engage in some silent reading.

Manner of Assessment. The reading processes being assessed are the mental operations which take place in the reader's mind during reading, in the form of a series of successive and simultaneous (Das, et al., 1979) organizations and reorganizations (Goodman and Burke, 1972). During the reading act these processes are not observable or generally open to introspection; therefore these processes have to be inferred from some form of reading product. Samples of both oral and silent reading may be used for this purpose. As the child reads orally, information concerning processing may be gleaned utilizing the techniques of "miscue analysis" (Goodman and Burke, 1972). This system of analysis however, is not possible during silent reading, as some manner of post facto inferral of processes is required. Insights into processing other aspects of silent reading

comprehension have been gained through methods such as: question answering, oral or written recall, and introspection.

In addition to oral reading errors an unaided oral recall following oral and silent reading will be used as product suitable for inferring process. As a consequence, this assessment will take the form of an informal reading inventory where continuous prose is read both orally and silently and two types of reading product are available: oral reading and unaided recalls. The choice automatically sets one obvious restriction on the assessment situation, namely, assessment must be individualized. The work of two groups of theorists--characterized by Beebe (1981) as the cue selection theorists and the discourse processing theorists--will be used to construct a theoretical basis for the two types of analyses.

These theories have developed relatively independently. A study carried out by Beebe (1981) attempted to reconcile these two theories and ascertain if they yielded complementary or competing accounts of reading comprehenison. One of the questions which the study tried to answer was: Are the units of analysis identified by both theories accounting for similar types of differences in reader's comprehension of prose as measured by a standardized reading test? These theories were found to be complementary: "at least some of the components of each theory were important predictors of comprehension when the theories were combined" (Beebe, 1981, p. 195). The basic tenets and assumptions of these analyses will be outlined below.

Miscue Analysis: Since the mid-sixties a series of

investigations into reading as a psycholinguistic process has been carried out "under the general direction of Kenneth S. Goodman" (Goodman and Burke, 1972, p. 9). These studies are simultaneously exploring the nature of the reading process (e.g., Beebe, 1980; Biemiller, 1977; and Weber, 1970) and refining the measurement of this process (e.g., Haupt and Goldsmith, 1982; and Hood, 1976). All these studies explore reading through an anlysis of oral reading errors or "miscues" (as Goodman and his followers prefer to call them). In fact, comprehension is the overriding criterion of reading in a miscue analysis. "Reading is not reading unless there is some degree of comprehension" (Goodman, 1968, p. 26).

Miscues, and not exact renditions of the text, are considered for investigation because "When expected and observed responses match, we get little insight into this process" (Goodman and Goodman, 1977, p. 319). The nature and quality of this mismatch between the written words and what the reader actually says lies at the heart of miscue analysis. An explanation of why these mismatches or miscues occur rests on four main assumptions. According to Goodman and Burke (1972, pp. 10-15) these are;

1. All readers come to the reading situation competent in their native language.
2. Likewise, readers bring their conceptual framework or store of world knowledge.
3. The text, however, represents both the language patterns,

interests and experiences of the writer.

4. Reading is not the passive reproduction of the words on the page; it is an active language process which involves constant interaction between the reader and text.

From these assumptions it follows that deviations will occur whenever there is a mismatch between "the language of the reader and the language of the author" (p. 5). Furthermore, it is not the quantity of miscues that is important but the quality. "All miscues are not 'equal' because some retain grammatical and semantic correctness and therefore detract little from the comprehension while others distort meaning considerably" (Beebe, 1981, p. 40).

Comprehension occurs when the reader selects certain cues from the page of print and integrates them with his prior knowledge which includes knowledge of the meaning system and grammar of the language. As a result of this knowledge, readers can anticipate (Beebe, 1981) or predict (Smith, 1975) much of the printed message. In addition, different readers (e.g., beginning vs. mature readers) possess varying degrees of knowledge about spelling patterns and grapho-phonic relationships which aids in the efficient selection of graphic cues. But generally a reader has prior expectations concerning the syntactic and semantic structures that he will read. So as he reads he selects a minimum of graphic cues in accordance with these expectations. However, if he is unable to associate these cues with his language or world knowledge, he regresses and picks up more graphic

information correcting his predictions, if necessary. But if, as usually happens when a reader reads familiar material, the cue selection is successful, he associates the graphics selected with his linguistic structures, synthesizes the meaning and proceeds to the next unit of text. This set of interactive processes is what Goodman (1976) calls the "psycholinguistic guessing game".

The instrument used to elucidate the pscyholinguistic game played between reader and text is the Reading Miscue Inventory (or some variation of it; see Hood, 1976). The strategies or processes in this game are observed through qualitative analysis of errors. The miscue inventory provides the researcher with a series of questions focusing on: the semantic, syntactic, and grapho-phonic similarity of the error with the text. These questions attempt to highlight the reader's use of the various cueing strategies. Two further questions basic to the cue selection theory outlined above are also posed: Is the miscue corrected? and Does the miscue result in a change of the author's meaning? Such an analysis of oral reading errors reveals that errors are not random; errors reveal a pattern which reflect the reader's hypothesizing. Hence Goodman's insistence on the label of miscue rather than error: he maintains that both the exact responses and miscues are manifestations of the same psycholinguistic process. This is why the majority of errors made by a proficient reader are meaningful.

In order to utilize the information available through miscue analysis the reader must make a number of errors. As a consequence the reader must read a passage that is slightly too difficult for him. In the case of school children this is achieved by asking the children to read passages one or two grade levels above their level of performance.

Recall Analysis: Although miscue analysis is an indepth qualitative analysis of oral reading errors, it operates specifically at the level of the sentence as defined in the transformational grammar (Chomsky and Halle, 1968). On the other hand, an unaided recall of the contents of a passage yields insights into how the entire passage is being processed via memory: "retelling provides an overt measure of what a student has done with the information presented in a selection, indicating that it has not been assimilated into existing schemata or scripts" (Pearson and Johnson, 1978, p. 129). This type of an analysis of recall--analysing a recall by comparing it with the original text--allows a researcher to view some of the "higher-order organizational skills utilized during the interpretation of the text" (Beebe, 1981, p. 54).

The work of the discourse processing theorists will be utilized to detail the comprehension processes that can be inferred from recall protocols and even more specifically for this investigation these theorists provide systems for analysing protocols so that comprehension processes can be inferred.

The study of recall analysis was expedited in the mid-1970's with the fusing of knowledge gained from two areas of study:

cognitive schema theory and text grammars. Recall analysis is based on certain assumptions about the schema or cognitive organization of the reader's knowledge and the structure of the reading passage. Perhaps the main assumptions may be characterized as follows:

1. Every reader has a hierarchially organized store of prior knowledge; hence, he instantiates what he reads by slotting information into existing structures (see following section on reader's knowledge).
2. Information on the page is usually highly structured and suggests the author's schema.
3. The structures of certain passages are very conventional and so are known to the reader and author alike, e.g., the structure of simple stories (Mandler and Johnson, 1977) or research reports (Kintsch and Van Dijk, 1978).

As a consequence of the assumptions listed above, it is posited that there is an active interaction between the structure and content of the text and the reader's knowledge and expectations, as the reader abstracts the information essential to convey the message. Thus, the recall will not consist of a random collection of some of the text information, but when mismatches occur between the author's schema and the reader's schema, recall will not be entirely faithful to the original text either in content or structure (see Bartlett, 1932; and Steffensen, et al., 1979).

The explanation of text comprehension proposed by the discourse processing theorists has many similarities with the theory underlying miscue analysis. Two of the differences, however, are (1) the time at which the reading product is observed, and (2) the focus of the recall analysis on the integration of text information as a whole. And yet, to understand a passage, subparts must be processed at some stage, because "a discourse is processed as a multi-level structure containing units as 'small' as individual concepts and relating connecting concepts, and as 'large' as macro-structures (total structure) consisting of connected propositions" (Frederiksen, 1977, p. 53).

Unlike miscue analysis where all research is based on a single paradigm (Kuhn, 1970), several different systems for analysis exist, each emphasizing a different aspect of text and recall organization reflecting the particular biases of its specific theoretical bases. Some of these systems are detailed below:

1. Story grammar focuses on an analysis of the macro-structure or overall structure of the text or recall. The basic units of comparison are story propositions. They correspond directly to the surface expression of the story and may be defined structurally as t-units. These units, however, are then parsed in terms of how they advance the underlying storyline.
2. Propositions were specified by Kintsch and his

colleagues in order to analyse texts. A proposition or micro-unit is defined as a predicate with all its arguments (nouns, adverbs, or adjectives)(see Turner and Greene, 1977 for a detailed account). These base units are elaborated into coherence networks to account for the structure of texts and recalls. Gist propositional similarity is the criterion for judging the match between recall and stimulus passage.

3. Another set of researchers began to examine the different types of text information recalled by assigning that information to recall categories (Drum and Lantaff, 1977; and Clarke, 1981).

The comprehension categories of Fagan (in press) which will be employed in this study to analyse the oral recalls are based on the Drum and Lantaff (1977) categories. Fagan, however, has revised these categories in accordance with the work of Kintsch and Van Dijk (1978), so that now they are suitable for inferring processes; each category is based on certain assumptions about the underlying processes that may have contributed to the kind of information coded in a particular category.

Elapsed Time: By necessity a recall must be made when reading of the passage is complete; exactly how long afterwards is a decision made by each researcher. Some researchers have asked the recall immediately following the reading (e.g., Forester, 1978), while others have waited several months (e.g., Bartlett,

1932).

The time that elapses between reading and recall has implications for the type and organization of information contained in recall protocols. While it is generally agreed that memory is both a constructive and a reconstructive process, some abstractive processes are also involved (De Beaugrande, 1981). In general, the longer the time between the reception of information and the recall, the more the reconstructive processes that may be involved (Bartlett, 1932; Kintsch et al., 1978; and Mandler and Johnson, 1977); or in other terms, the less information that is recalled in a verbatim-abstractive fashion, the more it is organized according to the schema of the reader. For this reason, many recent studies have engaged the reader in an interpolated task between reading and recall (e.g., Tierney et al., 1978).

An assessment of reading aimed at gaining information on how text information is assimilated and accommodated into existing schemata should not engage the reader in an immediate recall; at least five minutes should be allowed to elapse between reading and recall.

Situational Variables. The sample of reading behavior observed in the assessment situation may be influenced by factors within that situation. Because of the decisions which were reached concerning the mode of reading and manner of assessment, the examinee will be in an individualized situation. Even so, this situation can range from a routine in-class periodic assess-

ment of reading to one where the child is taken to a strange environment to have his reading difficulties diagnosed by a reading specialist. Then it would also seem that the examinee's understanding of the purpose of the assessment may influence his emotional state during testing, in addition to stress and anxiety caused by unfamiliar tasks and threatening environments. These influences should be lessened by having the student first read and recall a practice passage, at least, it should insure that all examinees know what is expected of them.

As this is an individual assessment, the sex, race, age and interactional style of the examiner may play a crucial role in the quality and quantity which are yielded. Therefore, steps should be taken to ensure a good rapport with the examiner and the examinee. This assessment will also be influenced by directions given by the examiner, both before the reading and the recalls. If the examinee accepts the context as valid, these directions will also determine his purpose for reading.

The Reader

The reader is an important component of any reading situation. In order to read, the reader must possess certain knowledge. This knowledge base can be divided into three components; knowledge of the world, knowledge of the language, and knowledge of the reading task.

Knowledge of the world. Anyone who has lived has amassed a store of knowledge about the world. All children who attend school in North America today have certain experiences in common. All will have had contact with certain institutions. A large pro-

portion will watch at least some television each week and will also have knowledge about friends, teachers, parents or guardians, siblings, certain games, animals, cars and buses, etc.

From his earliest days the child abstracts from each situation or communication the features that are most pertinent to him. And so he stores a succession of these abstractions, and constructs an internal representation of reality. These representations stored in memory become the raw materials of anticipations of other events and objects. Various theories of how knowledge is stored in memory, such as the network or the set theoretic, have been put forward (see Cohen, 1979). But for the purposes of this study, Minsky's (1975) notion of frames is useful. Frames are hierarchical structures of facts or event sequences with a certain amount of inbuilt flexibility to assimilate and accommodate new information; also lower level frames can be linked together into higher level frames when a new understanding or synthesis is reached. A network of frames forms a cognitive structure--an organized body of concepts.

A cognitive framework is operationalized for reading comprehension and memory by invoking the notion of schema (also referred to as script, or plan)(e.g., Schank and Abelson, 1977). A schema is a set of expectations that provides a perceptual set so that only certain enviromental information is systematically attended to. This selection also operates in memory reconstruction (Bartlett, 1932).

Schemas exist in the mind for anything with which one is familiar, from a friend's face to the consequences of certain

events or the structure of familiar stories. Therefore, schema can apply both to the content and structure of what is read. A thoroughly researched use of the notion of schema is that of "story schema" which is used by many authors to refer to a set of expectations about the internal structure of stories which facilitate both story comprehension and recall (e.g., Mandler and Johnson, 1977; Rummelhart, 1975; and Stein, 1978).

Simple stories have a very definite structure. The main protagonist is introduced, acquires a goal, attempts to achieve it, this attempt must have a result either positive or negative. Story schemas are idealizations of that structure which is internalized by listening to many stories and by participating in goal oriented action sequences since early childhood. Research carried out with school children from grades one to six show that they have a well developed story schema (Stein, 1978); younger children recall less but in exactly the same sequence as adults and older children (Mandler and Johnson, 1977). It can be assumed therefore, that all elementary school children will be familiar with passages following a story structure.

Knowledge of Language. Many aspects of linguistic development may be considered to affect how the child comprehends a text. In the present study, however, the focus will be on the syntactic component as this facet will be utilized to analyse the passages read in the present assessment.

A study carried out by Hunt (1965) of the written syntax

of students from grades four, eight, and twelve indicated that the younger students produce many short t-units (less than nine words), but with age there is an increase in the length of the main clause plus an increase in the number of subordinate clauses produced. These findings were confirmed by O'Donnell et al., (1967) for the oral language of kindergarten and elementary school children. Both studies conclude that the mean length of t-unit is a sensitive measure of development towards maturity in children's language production. Since there is little information available about the receptive abilities of children, it is assumed that they will also be able to read t-units of greater syntactic complexity as they get older.

Children's vocabulary also increases over the years and from nine to eleven years many hierarchical semantic distinctions are established (Clark and Clark, 1977).

Knowledge of the Task of Reading. The child must have certain conceptual and practical knowledge concerning the tasks of reading and assessment or testing. The reading situation must be approached with a tacit knowledge of the mechanics of print, purposes for reading, and with the expectation of understanding what is read. Whereas an assessment situation may be very alien and threatening to some students, others may be thoroughly familiar and at ease in such a situation.

The Text

Another basic component of the reading situation is the text. In this section, text will be conceptually analysed into the following components: print, vocabulary, syntax, structure, and

topic--thus emphasizing the hierarchical nature of texts.

Print. Print is the medium for conveying the message. It is a visual display comprised of twenty-six letters and seven common punctuation marks which are sequenced into words, and the words into sentences.

The English orthography is classified as an alphabetic system although many of the sound-symbol relations are far from perfect. Chomsky and Halle (1968) claim that the system is morphophonemic rather than alphabetic. In spite of this irregularity, many of the spelling patterns are highly predictable.

Vocabulary. Vocabulary refers to both the lexical items and their underlying concepts. The vocabulary of written language is often more formal and abstract than that of oral language (Rubin, 1978). Many lists of the words used in written language exist but the collection made by Carroll et al.(1972) is extremely useful for describing the vocabulary used in materials written for American elementary school children. In addition this work supplies many useful indices of the frequency and dispersion.

Another type of vocabulary frequency to be considered is the number of times the same words are used within a passage (Goodman, 1981).

Syntax. Syntax is the primary means used in language to convey the intended relation between words. Word order is the most important way of establishing these relationships in English. Syntax describes the structuring of the text at the t-unit level (Hunt, 1965).

Although very similar, the syntaxes of oral and written

language are not identical. In general the syntax of written language is more formal and complex. The average length of t-units is often used as a measure of syntactic complexity (Hunt, 1965; and O'Donnell, et al., 1967).

Conventional readability measures are usually concerned with calculating the average sentence length.

Structure. Text structure refers to two things: the organization of the underlying semantic propositions (Fagan, 1978a; Frederiksen, 1977; and Kintsch and Van Dijk, 1978) and the structuring of the content as described by the text grammars (Bower, 1976; Mandler and Johnson, 1977; Rummelhart, 1975; and Stein, 1978).

According to Kintsch a passage is comprised of a hierarchical network of semantic propositions; some propositions are considered higher level or macro-propositions. A network of these macro-propositions produces a gist or summary of the text. For a passage to be readable in Kintsch's terms it has to have an obvious macro-structure and repeated arguments (concepts); otherwise the coherence of the passage breaks down.

The system of text analysis developed by Fagan (in press) is similar to Kintsch's analysis. But unlike Kintsch who bypasses the syntactic system, Fagan has identified syntactic units which are in a one to one relationship with the semantic propositions which they represent.

The structure of the text described by the text grammarians is along similar lines to the underlying semantic organ-

ization described above, but that different base units and different text relations are employed. A story is an account of a goal-oriented action sequence and generally follows a predictable sequence: setting plus one or more episodes plus ending. Furthermore, the structure of a typical story on the page supposedly mirrors the schema in the mind of the reader.

A story grammar describes the hierarchical relations between the story components and supplies a set of generative rules to facilitate the writing of well-formed stories.

Topic. The topic is what the text is about. Usually the title is a pointer to the topic and in many narrative or descriptive writings it acts as an advance organizer.

Rubin (1978) indicates that the topics of written language often differ from those of oral language with which the child is familiar. While the latter topics deal with everyday objects and situations arising from the shared knowledge base of speaker and listener, the topics of written language may deal with abstract or unfamiliar subjects. Therefore the writer and reader may not be operating from a shared knowledge base and the writer has an incomplete model of the reader. So as a result, the subject matter of the text may be outside the reader's store of world knowledge.

Interaction Between Reader and Text: Reading Processes

The theories underlying the two methods of assessment outlined in previous sections depict reading as a constant

trade-off of information between reader and text, and conceptualize this exchange as a set of cognitive processes which occur during and after reading. Because the manner of assessment involves inferring process from oral reading and recall protocols, two sets of processes are involved: the processes involved in the reception of information gained from text and the processes involved in the production of reading product.

The reception processes are those processes that occur when the reader reads, that is, interacts with print and understands the author's message. These processes probably occur concurrently with oral reading and can be inferred from an analysis of oral reading errors, hence extensive use will be made of psycholinguistic theory in detailing the processes that can be inferred. On the other hand, the production processes are those processes which occur when the reader retrieves the text information from memory, and constructs and organizes it to produce an unaided recall which is a text in its own right and not simply "a replica of a memory representation of the original discourse" (Kintsch and Van Dijk, 1978, p. 374). A recall protocol is then the result of both receptive and productive processes.

But both sets of processes--receptive and productive--are facilitated and shaped by certain memorial processes. In very general terms, it may be asserted that constructive memory mediates receptive reading comprehension, with reconstructive memory mediating the productive processes. Pushing this temporal comparison even further, it is likely that constructive memory

is at work during oral reading, but reconstructive memory plays a greater role in the shaping of an unaided recall especially when time has elapsed between the reading and subsequent recall. Oftentimes researchers espouse either a constructive (e.g., Ortony and Anderson, 1975) or a reconstructive (e.g., Bartlett, 1932; and Spiro, 1977) approach to the conceptualization of memory. De Beaugrande (1981) shows the assumptions of these approaches. "The constructive approach assumes that the processor begins integrating stored knowledge with the presentation rightaway, so that memory receives an expanded, modified version of the experience and presents this when recall is required. In the reconstruction approach, further contributions are still entering after the experience is stored in memory; recall is thus based on the current state of storage being assembled by means of a general organizational pattern" (p. 264). Clearly, length of time elapsed between the recall and the experience is a crucial factor in determining which type of processing is most likely to occur.

The question of what type of memory is operating is not one of "either-or", but of degrees of influence. Thus, in addition to a combination of constructive and reconstructive memory processes, an unaided recall may also be influenced by "abstractive" processing (Tierney et al., 1978) which in turn may have been influenced by complementary reception processes. For example, Kintsch and Van Dijk (1978) explicate three reception processes involved in the formation of a gist and comprehension of a complete text: (1) deletion of irrelevant information,

(2) generalization of sub-sets of information, and (3) construction of global facts from specific information. These authors posit three complementary processes for retrieval of information from memory to produce a recall protocol. The first of these is "reproduction" which is involved in the recall of information stored in a verbatim fashion. "Transformation" is involved in reordering text information, paraphrase, explication of coherence relations, and perspective changes. The third process "reconstruction" effects an interaction between text data and world knowledge resulting in (a) the addition of plausible details and normal properties, (b) particularization of events, and (c) specification of normal conditions and components.

Therefore in the Kintsch and Van Dijk system of analysis, recalls are assumed to be the product of comprehension and memory processes. Even though some researchers (e.g., Fitzgerald Whaley, 1981) talk about eliminating or minimizing the effects of memory, according to either cue selection theory or discourse processing theory reading is impossible without memory; even more generally, comprehension or perception is impossible without memory or more precisely, semantic memory (Shoben, 1980). This confusion may be clarified by distinguishing between "use of memory" in comprehension, and "memory for" that comprehended event (episodic memory).

Another point of note is the distinction between <u>meaning</u> of the author, <u>comprehension</u> of the reader, and memory.

In a normal reading situation a reader focuses on comprehension and not on memory. Memory may be considered "the automatic or involuntary product of comprehension" (Brown, 1978, p. 107). A similar point is raised by Clark and Clark (1977) who ask: "Is memory in normal situations really memory for 'meaning'? Not exactly. It is memory for the products of comprehension, and they aren't necessarily the same as the meaning of a sentence" (p. 153), that is, memory concerns the situation or context surrounding the meaning in addition to the actual meaning. This, however, is how the reader understood the meaning; hence, it can be argued that tasks involving memory do measure comprehension.

At present there is no way of knowing whether the schema which controls reception is the same as the schema used in production of a recall. But it may be assumed that if a reader deliberated about what he had read, he would engage in processes similar to those that occur during the production of information. Hence, this research will not explicitly attempt to distinguish between the processes which occurred during reading or after the reading is completed. In conclusion, one may argue that the point from which the reader connects with print to the point at which he demonstrates his understanding of it constitutes the reading act. In turn this study may be considered to be focusing on the processes the reader engages in during the reading act.

Nature of Reading Processes. The processes employed in cognition in general are highly similar or identical to the processes of reading, general mental processes such as attending, analysing, categorizing, etc. In this investigation, nevertheless, an effort will be made to define and delimit each process in terms of its role in both the word identification and comprehension stages of reading. Not every process is employed to the same extent or at each stage in the process of bringing meaning to print and print to meaning. Furthermore, some of the processes may be considered as being more "data-based" or "bottom-up", while others may be predominantly "conceptually driven" or "top-down" influences (Norman and Bobrow, 1976). Reading comprehension is then the result of an interaction between these two sets of processess (Rumelhart, 1977).

It should be noted, however, that the reading processes which have been selected for this assessment do not represent an exhaustive account of reading. For example, categorizing (Rawson, 1979) is an important reading process, but due to the technical limitations of the present miscue and recall analysis, it is not possible to infer it from either the oral reading or recalls. This study attempts to assess the reader's engagement in the processes developed below.

Attending. In western psychology and philosophy, attention and consciousness have long been discussed as complementary phenomena. The following definition, given by William James in 1890, captures the essence of this relationship:

> It is the taking possesion by the mind, in clear and vivid form, of one of what seem several simultaneously possible objects or trains of thought. Focalization, concentration of consciousness are of its essence (In Lupart, 1981, p. 13).

The process of attending determines the amount and type of external stimuli which reach the mind through the various sensory channels. Allington (1975) has characterized attending behavior in the classroom as "looking at or listening to the features of a stimulus which will provide the basis for an appropriate response" (p. 22). Embedded in this definition is the notion of purpose. Purpose, but also knowledge, helps to explain selective attention. When trying to find his way in a strange city, a driver does not attend to the hundreds of billboards, neon lights, and store names, but selects "useful information"--highway and street signs--for attention.

Precepts may enter into consciousness (sensory store) through all the sensory channels. A listener may attend simultaneously to the voice, gestures, and posture of the speaker; whereas the reader is provided with only visual stimuli. This probably explains why attention has generally been of interest to reading theorists who hold a bottom-up view of reading. These theorists have put heavy emphasis on the perception and utilization of graphic information in explaining reading comprehension. An extreme example is the model proposed by Gough (1972), where the reader is expected to attend to each letter of each and every word. The models of reading as a successive series of feature extraction, presuppose attention to word features such as spelling

pattern; thus demanding high reader "utilization" of presented graphics (De Beaugrande, 1981).

On the other hand, some top-down theorists see the visual scanning system, reader purpose and background knowledge, as of primary importance in directing attending behaviour. With the aid of scanning strategies, peripheral vision, and predictions, the eyes attend only to the appropriate visual cues (Hochberg, 1970; and Goodman, 1975). Readers of each language must develop appropriate scanning and attending strategies. "Written English progresses from left to right. Consequently a reader must develop scanning strategies which respond to this characteristic of language" (Goodman, 1973, p. 60); and thus learn to ignore the "irrelevant" (Gibson and Levin, 1975), or the "redundant" stimuli (Smith, 1978).

Dodd and White (1980) argue that any theory of attention must explain two phenomena: (1) ability to attend selectively, and (2) divided attention. Both issues have been raised above when discussing top-down and bottom-up views of attention. In the latter, bottom-up theory, a reader must pay almost constant attention to graphic input; as a consequence he has little spare processing capacity for meaning association and synthesis. The beginning reader who plods through the graphics is constantly plagued with the problem of overload on decoding; thus he never gets to meaning. Samuels, himself a predominantly bottom-up theorist (e.g., Laberge and Samuels, 1974), offers a solution to this overload dilemma, and he stresses the need for automatic

attention: "behavior is automatic when it can be performed without attention" (Samuels, 1976, p. 324). Smith (1978) refers to this overload problem as one of "tunnel vision": the eye attempts to attend to all the visual stimuli but the brain is only capable of processing a small proportion of those at any one time. Unlike Samuel's solution, Smith emphasizes the need for selective processing of the visual information in accordance with prior semantic and syntactic predictions. Whereas Samuels would claim that a reader may divide his attention between attention to grapho-phonics and comprehension when the lower level process of attending can be executed automatically, Smith on the other hand, would claim that divided attention is the result of selective attending, which in turn results from higher level processor contributions.

A bottom-up model of reading cannot answer the problem of selective attention because such a model allows a minimum role to processor contributions (see De Beaugrande's (1981) commentary on the Gibsonian model pp. 272-273), and reading was defined as a consistent interaction between reader and text. Therefore, a completely top-down model is also not a valid solution, as reading in this manner would not require attention to the graphics or an exchange of reader and text information, but rather an arbitrary imposition of meaning on text. The interactive view allows for flexibility in attention, and gives the reader autonomy over his scanning stategies, allowing him to use graphic information selectively according to purpose.

Attending is the first step the reader must make in order to decode print. In this model, attending refers only to graphic attention or focusing attention on graphic information; it is not conceptualized as the reader's over-all attitude toward a particular reading task (Lupart, 1981).

Analysing. Whereas attending is viewed as a relatively passive force in word identification, analysing is an active organizing force, and is occupied in selecting visual information by the discrimination of its distinctive features, and then abstracting and organizing this information for further processing (Malicky, 1982). In reading it may be viewed as the ongoing structuring of visual information from the instant the image hits the retina until recognition is made. At present there is renewed interest in explaining how humans recognize patterns in cognitive domains such as music, writing and print styles, abstract codes which people use in various occupations, etc. Recently it has been established that humans possess specialized feature detector cells (Dodd and White, 1980) which respond to certain physical characteristics of the stimuli, e.g., pitch, or wavelength of color, or shape. More specifically for the detection and abstraction of visual information, Gibson asserts: "There is evidence at a neurophysiological level that straight lines, curved lines, horizontals, verticals and diagonals are abstracted and responded to as features of patterns" (1969, p. 88).

Much of the early work of Gibson was devoted to discovering the distinctive features of letters and to studying the

development of children's ability to discriminate letters (1962, 1970). She explored the perception of features like "rotation and reversal" (b, d, p, q), line-to-curve" (u, v), and "break-close" (O, C). Despite the overriding emphasis on perception, she concluded that letter perception problems have their roots in cognition: the child can "see" the difference, but does not realize that it is a significant feature for the discrimination purpose at hand. In her more recent work, however, she proposes spelling patterns as the crucial units of analysis in word recognition (Gibson, 1976).

Many models of reading specify a feature analyser or detector to analyse and categorize visual information, although there is little agreement about the units of analysis. In Laberge and Samuels (1974) model, the visual icons are scanned by "feature detectors" which extract information about angles, lines, intersections, etc. Eventually these features will be fed into letter codes, or directly mapped onto spelling pattern codes or visual word patterns. Unlike Gough's (1972) model, which maintains that every letter must be analysed, Laberge and Samuels' model allows for the possiblity of larger units of analysis. Rumelhart's (1977) interactive model posits a "feature extraction device" which operates on the visual input, extracting the critical features and which will feed sensory input to "the pattern synthesizer". But analysing is greatly expedited by both "top-down" (semantic and syntactic) and "bottom-up" (letter and spelling patterns) hypotheses. Furthermore, while this processing is taking place

lexical hypotheses are being formed as to the identity of the word.

Visual discrimination is an important component of Smith's (1978) theory of reading, and he devotes much space to speculating on both the features of words and letters: "A distinctive feature...is an element of a stimulus configuration that constitutes a 'significant difference'-- that enables a perceiver to eliminate some of the alternative categories to which a configuration might be allocated" (1978, p. 132). However, he introduced the notion of "redundancy" (information that is available from more than one source while reading) to explain why the reader analyses different units and quantities of print, depending on his prior knowledge about what the letter, word, or phrase might be. On occasions the reader is assumed to recognize some larger meaning units directly, with a minimum of analysis. However, there is a definite limit to the extent to which this may occur, because of the linear nature of print.

Although there is disagreement as to the basic unit of analysis, all theorists agree that some form of analysis must take place because it is impossible for the human eye to perceive a block of print and abstract its meaning. In this study, analysis refers only to the discrimination, organization, and abstraction from visual information supplied by print. It is further assumed that this process is influenced by purpose, knowledge, and predictions. For example, where the reader's purpose is to verify every word on the page, as in proof reading, he must analyse the graphics carefully; a similar situation

occurs when reading a technical manual in an area where the reader possesses little background information. However, the accessibility of the reader's memory store of knowledge about letter and word features will also determine the operation of analysis.

Sequencing. In this study, sequencing is viewed as an adjunct process of attending-analysing. Because of the linear nature of print, letters and words must be processed in an overall left-right direction. Nevertheless, many words in English cannot be pronounced if read in a strict left to right succession. A final 'e', for example, signals a certain pronunciation of the preceding vowel; so that in spite of the general left-right scanning of print the reader "must accommodate his use of scanning strategies and the information they provide, so that he anticipates or seeks out information which may not be in a simple left to right sequence" (Goodman, 1973, p. 60).

Sequencing problems may arise at the letter or word levels. The single distinguishing feature of many words is letter sequence, e.g., bran for barn or clam for calm; thus, failure to sequence letters may result in comprehension problems. On the other hand, although word sequence plays an important role in conveying meaning, the same or similar meaning can result from different word sequences, e.g., "Mary says" instead of "Says Mary". These latter sequencing miscues may result from semantic or syntactic prediction, especially if the language patterns of the author deviate from those of the reader. The ability to use syntax obviously

rests on the ability to sequence.

In addition to sequencing graphic information, the reader must also abstract and store the semantic content of a text in the appropriate order. Story grammar, for example, analyses a recall not only in terms of its content but also in terms of the sequencing of that content in comparison to the stimulus story. This type of analysis, measuring the sequencing of recall information, is not proposed in the present investigation.

<u>Associating</u>. Association is virtually the whole underpinning of behaviorist theories of learning. Contiguity conditioning--whether through strengthening stimulus response bonds or operant conditioning--was seen as the basic mechanism of learning by many North American educators, from the work of Watson at the beginning of this century until the Skinnerian programmed learning in the late 1950's and 1960's. It is hardly suprising, then, that many of the "great debates" in the teaching of beginning reading during the same period concerned "units of association": whole words with word label, letter patterns with speech patterns, or single letters and digraphs with phonemes. Learning to read was simply a matter of association: "The great task of learning to read...consists in learning the very abstract equation: <u>printed letter = speech sound to be spoken</u>" (Bloomfield, 1942, p. 42).

The teaching of phonics is based on an assumption similar to Bloomfield's equation; namely, one of the crucial

processes of reading involves decoding letters into speech sounds, and in this manner, written to oral language. "While many writers agree that this association is necessary for beginning readers, there is considerable controversy regarding the use of phonological recoding by proficient readers" (Malicky, 1982, p. 2-12). This moot point has been taken up by Smith (1978) and Goodman (1976), who hypothesize that the proficient reader bypasses phonological encoding, associating print to meaning, while the beginning reader has to go through the intervening auditory or oral step.

Therefore, two types of association may be involved in reading: meaning and sound-symbol association. Smith (1978) refers to the former as "immediate word identification", where the graphics immediately trigger off the meaning of the word. To engage in this kind of association the reader may have a store of "sight words" and their meanings encoded in memory, so that perception of the graphics immediately prompts a semantic association. On the other hand the proficient reader, who relies heavily on higher level processing, need not associate every word on the page in order to construct its meaning. To engage in sound-symbol association, the reader must utilize a more specific knowledge store: knowledge of grapheme-morpheme relationships, or spelling pattern-morpheme or spelling pattern-sound unit. Proficient readers engage in this strategy whenever they read a new word which they cannot recognize with the aid of context. Smith (1978) refers to this strategy of using a combination of sound-

symbol and meaning association as "mediated word identification".

Association also plays a role in the storage and retrieval of larger units of text such as phrase or clause. As the reader reads he maps the meanings of the units of print onto his own semantic structures. These units may be stored in a verbatim manner and "reproduced" with the appropriate cue, or the gist meaning may be abstracted and the unit may be recalled in a similar manner but "transformed" (Kintsch and Van Dijk, 1978). Two categories of the recall analysis attempt to measure this facet of association.

In the present theory, association whether of sounds or meaning is conceptualized as an automatic process of a stimulus-response type. This process is however aided by higher level processes such as meaning synthesizing or the use of context (Shoben, 1980). It is also constrained by the units which have been analysed and attended to.

Predicting. Predictions are always made on the basis of knowledge. The daily weather forecast is made by meteorologists who have analysed decades of amassed knowledge about similar atmospheric conditions. Likewise, scientists make and then test hypotheses about the behavior of phenomena in certain conditions. Thus, theorists who emphasize the use of background knowledge in reading necessarily highlight the role of prediction in reading. In fact the psycholinguistic theorists view it as the single most important process in reading (e.g., Goodman, 1969, 1976; Hochberg, 1970; and Smith, 1975, 1978). Other theorists like Rumelhart (1977)

and the test grammarians (e.g., Mandler and Johnson, 1977; and Stein, 1978) also grant a key role to hypothesizing in their model of reading and story comprehension. In the latter, a reader/listener's story schema operates as a set of expectations about the parts of an ideal story.

Prediction may occur at three levels during reading--at the orthographic, syntactic, and semantic levels. A detailed account of the mechanisms of orthographic prediction has been proposed by Smith (1975); in keeping with this account he refers to prediction as "the prior elimination of unlikely alternatives" (p. 306). Letters in words are in predictable patterns rather than a random sequence. Furthermore, some letters are far more common than others. As a consequence of these observations, people can guess the letters in a word such as stream in three tries if they are given feedback after each selection. As each successive letter is verified, the remaining choices are highly constrained, Smith (1975) offers an account of such an experiment (p. 307). If the reader did not rely on the "redundancy" inherent in print, reading would be slow and restrained.

Both syntactic and semantic prediction are the primary process of Goodman's conceptualization of reading as a psycholinguistic guessing game. The reader is an active language user and utilizes grapho-phonic information only to confirm his hypotheses. According to Goodman (1975): "The apparently proficient reader may in fact be due to more successful prediction and hypothesizing, rather than more careful use of visual information"

(p. 217). In partial confirmation of this assertion, Adams (1977) reviews research findings which demonstrate that the proficient reader is not only more sensitive to the constraints of syntactic structure and semantics, but also utilizes both to regulate the construction of meaning and perceptual processes (such as fixations). In addition, she reports on a study carried out by Marcel where readers were presented for only 200 milliseconds with strings of words. The readers were asked to report as much information as possible. The most suprising finding was that increased contextual constraint (given before the actual presentation) led to a "disproportionate" increase in the number of miscues which were simultaneously grammatically and graphically acceptable; this syntactic and semantic prediction "apparently increased the visual angle at which subjects could discern graphical details of the printed word information. This is a compelling demonstration of the interfacilitation between top-down and bottom-up processes" (Adams, 1977, p. 22).

Thus, higher level processes such as semantic or syntactic prediction can aid in the analysis of the graphics, but in turn the final construction of the author's meaning must be constrained by graphic analysis. The interaction between prediction and attention to grapho-phonic input highlights the exchange between the knowledge of the reader and the visual information.

In this model of assessment, prediction refers to the

simultaneous use of semantic and syntactic knowledge to anticipate sentence meaning and word order.

Monitoring. For sustained reading to take place, the processes of prediction and monitoring should operate in a reciprocal relationship. As the reader makes predictions, he has to confirm them instantly in terms of the presented graphic information, and his knowledge of word order and meaning.

> Having made predictions, the reader must use confirmation strategies to check on the consistency of his expectations with the cues he is encountering as he reads on. He must ask himself whether it is making sense and whether the grammatical pattern he has predicted is the one he is finding. The same graphic cues which he uses to make subsequent predictions are used to confirm or reject prior predictions (Goodman, 1973, p. 62, emphasis in the text).

If the reader is to realize that he has lost meaning, he must have brought the purpose and the process of reading to a certain level of awareness. Thus, in recent years the problem of monitoring has been studied as one of the meta-cognition (e.g., Brown, 1980; and Flavell, 1981). Although presenting reading as a psycholinguistic process, Ruddell (1976) depicts monitoring as a cognitive process involving evaluation. In addition to the awareness of loss of meaning, the reader must also judge when a regression to correct an error is worthwhile, e.g., for most purposes reading "serpent" for "snake" does not interfere with meaning. Furthermore, he must evaluate the effectiveness of his rereadings and monitorings, and the depth of his background knowledge.

In general, research based on miscue analysis shows that good readers tend to correct errors which detract from meaning, and that monitoring behavior differentiates good from poor readers from the initial stages of reading (e.g., Beebe, 1980; Clay, 1969). It must be noted, however, that at different stages in learning to read or as a result of the emphasis of programs of instruction, a reader might monitor himself in terms of different contexts. For instance, a longitudinal study carried out by Biemiller (1970) on the miscues of grade one readers showed that over the school-year, the children's errors were constrained first by the meaning, then by the graphics and finally the better readers began to use a combination of constraints.

Monitoring, in the present study, will refer to the correctional behavior the reader engages in during oral reading. (Questions concerning the quality of the corrected errors will not be addressed for the moment.) This method may not measure all the monitoring behavior a reader engages in. Oftentimes, monitoring does not take place immediately; a prediction is made, sampling further context before he realized his error: at this stage, he may simply make a mental correction.

Inferencing. Inferencing and predicting have been proposed as twin reading processes by Malicky (1982). This close relationship is probably based on the notion of "backward-" and "forward-inferencing" (Macleod, 1978). Backward-looking inferences are made as the reader links new information in the

text with previous information from the same text -- whereas forward-looking inferences are made using information read up to that point to predict future states or events. In the present work, inference refers only to the type of processing involved in "backward-inferencing", and operates as an agent of integration and cohesion. While reading, the reader instantiates the text with information by referencing it to his background knowledge or frames of reference for different states or events. Thus he is able to draw "implications" beyond those actually stated by the author (see Clark and Clark, 1977, pp. 154-156). Work by Johnson, Bransford, and Solomon (1973) shows how altering one word in a lengthy sentence may completely change the type of inference or implication made. In addition, they discovered that people experience difficulty in distinguishing the implied information from the stated information in a recognition test.

Some inferences are required to bridge elements in a passage and therefore introduces new information into this passage and the subsequent recall, e.g., somebody's reason for carrying out a certain course of action (Crothers, 1978). Research in story grammar has shown that older children make inferences in order to rationalize the irrational behavior of characters, Whereas younger children simply ignore nonsensical material (Stein and Glenn, 1977). Vague or unusual material is comprehended as it is given coherence in a familiar frame of reference; and inferences are made to comprehend certain states or events

(e.g., Anderson et al., 1977). In a recall following this type of reading, many reconstructive errors may be expected to occur, as the discourse content will have interacted with background knowledge — thus partly erasing the boundaries between new and old information.

Furthermore, in a recall an individual may consciously or unconsciously elaborate his association. Elaborating or embellishing may not be considered as an inference, as it is unsubstantiated by stated text material; but neither is it considered an error in comprehension or recall, as it does not contradict any of the explicit statements, being simply an embellishment of the schema introduced by the author.

Readers use different strategies to integrate text information as a coherent unit, thereby integrating this new information into memory. These inferencing strategies may separate proficient from non-proficient readers (Mcleod, 1978; and Phillips-Riggs, 1981). The type of knowledge a reader brings to the page and how he utilizes it to make specific textual connections in part explains how literary works are appreciated on different levels. This type of processing in which the written message is expanded in any way is viewed as reconstructive.

Synthesis. Whereas synthesis is often used to refer only to the reader or listener's ability to organize the various subparts of a discourse into a coherent structure, in the present work synthesis may operate at three levels: word, sentence, or

passage. Word level synthesis is involved in the blending of the phoneme, syllable or morpheme units into the corresponding word pronunciation. Sentence or phrase level synthesis results in the chunking of information contained in those text units for memory storage. Finally, passage level synthesis is involved in the abstraction and organization of discourse content to form a gist or macro-structure of the most important bits of information. The latter two categories of synthesis are part of the process called memory for substance by Clark and Clark (1977).

Based on the work of Henry (1974), passage level synthesis may be divided into two distinct sub-processes: (1) a verbatim summary, or synopsis, of the information on the page; and (2) a conceptual synthesis of the new information, where the reader integrates the new information into his existing knowledge base in such a manner as to create a true understanding of what is read. The first of these sub-processes is subsumed by Kintsch's processes of "transformation" of text data, and the second by the process of "reconstruction".

In order to synthesize the substance of any unit of text, the reader must become sensitive to some aids to synthesis within the discourse structure. For example texts are written on a consistent topic, thus once recognised, the theme will become the scaffold on which the reader will first construct and later reconstruct the author's meaning. As the coherence network is constructed, certain concepts and relationships will be sharpened while irrelevant or superfluous material will be discarded (Kintsch

and Van Dijk, 1978) as the reader utilizes his schema to synthesize the key concepts of the new information (Meyer, 1974). In addition, this information is structured within a predictable text structure such as story schema, within which the relationship between people and events are signalled by a set of referential, temporal, or logical connectives. The title of a passage acts as an advance organizer and is also a key factor in how it is synthesized and stored in memory (see Nicholson, 1977).

Research on story structure has shown that even very young children are sensitive to story structure and use it to organize incoming information (Brown, 1975). However, it is not until approximately grade six that they can cope with certain disruptions in the expected story sequence and theme (Mandler, 1978). Furthermore, research has shown that when the reader is familiar with both the structure and content of the passage, there are generally few synthesizing difficulties (e.g., Kintsch and Greene, 1978). On the other hand, if the structure or semantic content of the passage is alien, the reader will attempt to interpret the facts within a more familiar schema (e.g., Steffensen et al., 1978). So as a result the reader's synthesis will not mirror the intended message of the text.

Synthesis is the key process when committing text information to long term memory. It would be impossible to recall more than a disjointed sentence or two of a passage without some form of cross sentence synthesis, especially if time had elapsed since reading. The ability to give an oral recall where the

information is clearly organized and synthesized at a cross sentence level is highly correlated with reading comprehension (Beebe, 1981). As the sentences of the text are read, each sentence and phrase is synthesized and in turn this new information is subsumed into the global representation or coherence network. Sentence or partial sentence synthesis is aided by both syntactic and semantic prediction. Because the human mind is a limited capacity processor, new information must be encoded in long-term memory at various stages in longer sentences. Research results have demonstrated that this process normally occurs at syntactic boundaries (Fodor et al., 1974).

Interaction of the Processes. When reading is taking place, the various processes are inextricably intertwined. According to Goodman (1976), the reader simultaneously engages in all levels of processing on graphic, syntactic, and semantic information. Word identification and comprehension are instantaneous for some readers. In fact, Smith (1978) maintains that comprehension precedes perception; no process can operate in total isolation. However, for the purpose of assessing reading processes, single processes can be isolated and said to be dominant in a particular strategy of word identification or comprehension.

Limitation of the Present Theory

The focus of the present theory of assessment is on *how* the reader comprehends continuous discourse; in other words, the aim is to gain some insights into the processes a reader uses to get meaning from print. Thus, the focal point is the interaction

between the graphic information and various aspects of the knowledge of the reader. Prior knowledge is therefore crucial to the quality and extent of processing. Since this study will not measure background knowledge directly, attempts will be made to control both the structure and the content of the text.

It should be borne in mind that a theory is a transient product of the socio-historical context in which it originated. The criticisms directed at past theories may equally apply to the present theory at some point in the future.

Conclusion

The theory of reading assessment presented in this chapter will guide the construction and subsequent piloting of an instrument to assess reading processes. Reader will be interpreted as a Canadian elementary school child; text will be written on the topics of sports or animals and follow a particular story structure; and the assessment situation will include the researcher as the examiner, and include a training and testing situation. In the training component the child will be explained what is involved in the task, and will be given a practice story to read and recall.

The next chapter will introduce a theory of test construction appropriate for creating an assessment instrument to measure this complex construct of reading as dynamic cognitive and linguistic processing.

CHAPTER IV

TEST CONSTRUCTION

Overview: Scientific Knowledge and Measurement Error

> Science is a very human form of knowledge. We are always at the brink of the known, we always feel forward for what is to be hoped. Every judgement in science stands on the edge of error, and is personal. Science is a tribute to what we can know although we are fallible (Bronowski, 1973, p. 374).

All knowledge is made by people not by minds. No matter how abstract the conceptualization or model, scientific knowledge consists of manmade ideas. Scientific discoveries are simply what happens when inquisitive people try to see reality in a new light in order to better understand themselves and their world. Scientific knowledge is, then, a constant exchange between people and nature: the extraction of perceptual information through the imposition of reality sets, models and theories on natural phenomena. This knowledge is at best uncertain, limited and changing.

By the 1930's physics had lost its belief in the absolute significance of measures of objects and had yielded to the concept of a time-space continuum. The nature of the uncertainty is even greater in social sciences like psychology and educational measurement, because many of the phenomena from which knowledge is abstracted are not clearly defined as objects or events. Typically, the phenomena to be measured are based on abstract concepts such as: intelligence, ability, and reading comprehension. As a result, two questions immedi-

ately confront us: How good are our own conceptualizations of reality? and second, How well do measuring instruments capture the essence of these conceptualizations?

This problem of bridging the gap between theory and research is seen by Blalock (1968) as one of measurement error. This error is overcome, in part, by specifiying two qualities of all measurement, namely reliablity and validity. "Fundamentally, reliablity concerns the extent to which an experiment, test, or any measuring procedure yields the same results on repeated trials" (Carmines and Zeller, 1979, p. 12). The problem of reliability or random error of measurement was addressed by Gauss at the end of the eighteenth century. Gauss demonstrated that in the face of this type of uncertainty, the best estimate of a series of measures is the mean value (Bronowski, 1973).

In contrast, Carmines and Zeller (1979) judge validity a more fundamental quality than reliability; reliability focuses on certain properties of the empirical indicator whereas validity highlights the crucial relationship between the concept and its indicators, and may be said to deal with non-random or systematic error. Hence, theoretical claims are invariably made to validate a psychological measure, because it must be demonstrated that the measuring device is indeed measuring that abstract concept. This conceptualization of validity is called construct validity. And even more specifically in the case of a reading measure Kirsch and Guthrie (1980) state:

> Construct validity refers primarily to the psychological processes required to complete the items on a test. It

> pertains not to the kinds of information in the items, but to the cognitive operations needed to obtain the required information from the document. It refers not to the subject matter of the test content but to the relation between the knowledge structure of the examinee and that of the test. Construct validity is based on the psychological meaning of a test score and the theoretical explanation of a good or a poor performance on a test (p. 81).

Thus, construct validity is viewed as elaborating a process account of taking a test; however, because of the nature of scientific knowledge, the process of observation or test construction must also be unraveled in order to get at the significance of the measure.

It is the purpose of this chapter to unmesh some of these diverse strands of measurement in an attempt to devise a model of test construction capable of incorporating concerns of construct validation into all phases of test construction, and subsequent use. The thesis then being argued in this chapter is that constructing and using tests is construct validity; hence construct validity should be the overriding concern when constructing and using tests. This argument will be developed in three stages: (1) first the history of construct validity in psychometrics and its role in building scientific knowledge will be introduced; (2) second ways in which construct validity may be established will be explored; (3) finally, an existing model of test construction will be presented and then revised so as to award concerns of construct validation a central role in all decisions regarding test building and use.

In spite of its sometimes negative connotations, "testing" is simply "the means by which educators measure" (Maguire, 1981, p. 10-1). So for the remainder of this study "testing", "assessment", and "measurement" will be used interchangeably.

Construct Validity

History of Construct Validity in Psychometrics

The history of psychometrics is typically presented as the charting of various innovations in test construction, from the early work of Spearman through the synthesizing of classical test theory in the mid-1930's (e.g., Dubois, 1970). Construct validity is the most recent major conceptual innovation in the realm of validating psychological measurement. Although the notions underlying construct validity were adumbrated in publications in the early 1950's (e.g., Peak, 1953), it was not until 1954 that the broader view of validity was given official sanction. In that year the American Psychological Association published Technical Recommendations, a document which set forth standards for commercially available tests, listing the validational research which is required prior to distribution. Since validity was not conceptualized as a simple construct, four distinct types were named: content, concurrent, predictive, and construct validities.

Two other seminal works on the subject of construct validity appeared in 1955 and 1957. Cronbach and Meehl's (1955)

paper may be depicted as a sequel to the Technical Recommendations (1954), further explaining the additional concept of validity and elaborating its implications for measurement, and supplying a philosophical basis for the interpretation. But these authors are at pains to point out that they are "not in the least advocating construct validity as preferable to the other three kinds" (p. 74, emphasis in the text). Their main task is one of informing psychologists about the concept, "so that they can make a place for it in their methodological thinking, that its rationale, its scientific legitimacy and its dangers may become explicit and familiar" (p. 74). In contrast to the more measured stance of these two authors, Loevinger's (1957) monograph contains an all-out plea for construct validity, declaring in unequivocal terms that henceforth construct validity will subsume the other forms of validity plus reliablity. The notion of construct validity is expanded so that conceptually and methodologically all aspects of validity and reliability are embraced into the "substansive", "structural", and "external" components of construct validity.

Before that period validity was established by correlating the test scores with some observation that served as a criterion (concurrent and predictive validities). It logically followed that the primary aim of testing was to predict future performance, e.g., later scholastic achievement from performance on an IQ test. "The theory of prediction was very nearly the whole of validity until about 1950" (Crombach, 1971, p. 433). However,

in the intervening years, psychometricians and researchers have devoted increasing time to descriptive and explanatory interpretations of scores; and to developing rationales as to why a certain measurement was sought in the first place. As a consequence of its weak explanatory power, correlation or predictive validity has been assigned to a secondary position.

Since the pioneering days in the 1950's, construct validity has gained a firm foothold in the mainstream of educational and psychological measurement. But unfortunately this is too often more of a ritualistic than a practical concern both in testing and in research measures. Published tests in reading, for instance, until very recently failed to mention construct validity, and when discussed the supplied construct validity information is meager (see Wardrop et al., 1978). The antitest movement in the past decade and a half has forced psychometricians to be more aware of the qualities they are attempting to measure and the comparability of these measures when applied to different groups within each society (see Feuerstein, 1979, pp. 1-58 for a discussion of the anti-test movement).

To conclude this brief overview of the introduction and legitimization of construct validity the following quotation from Messick (1981) is given because it captures the essence of the change in attitude towards validation:

> In what at the time appeared to be a highly expansive moment, Loevinger (1957) proclaimed that: 'Since predictive, concurrent, and content validities are

> all essentially ad hoc, construct validity is the whole of validity from a scientific point of view'. In the 25 years since these words were published, it has become increasingly clear that this seemingly radical doctrine is in actuality a central principle of educational and psychological measurement and that, if anything, it does not go far enough in stressing the fundamental role of construct validity -- not just for scientific measurement but for applied measurement as well (p. 9).

The Logic of Construct Validation

> Construct validation takes place when an investigator believes that his instrument reflects a particular construct, to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim (Cronbach and Meehl, 1967, p. 65).

This understanding of the process of construct validation rests on a position or combination of positions within the philosophy of science (e.g., Feigl, 1953; Hempel, 1966; and Kaplan, 1964). In fact the logic of construct validation can draw on any philosophical stance other than those grounded in logical positivism. The advocates of the latter theory maintain that construct validity simply opens the door to fuzzy thinking and the obfuscation of psychological facts (Bechtoldt, 1967).

"Operationalists (logical positivists) argue that all useful concepts must originate from experience,...(and so) require operational definitions for all theoretical concepts" (Kemeny, 1959, p. 127). In order to adhere so strongly to this credence, they are compelled to believe in a model of scientific "progress" which begins in blissful ignorance and advances towards final truth by the successive accumulation of facts. But

"Historians of science have utterly discredited this model during the past decade... Changes in theory are not simply the derivative results of new discoveries but the work of creative imagination influenced by contemporary social and political forces" (Gould, 1973, p. 201). The work of Kuhn (1970) heralded this view of the growth of scientific knowledge:

> Todays's physics textbooks tell the student that light is photons...Research proceeds accordingly...That characterization of light is, however, scarsely half a century old. Before it was developed...physics texts taught that light was traverse wave motion...Nor was the wave theory the first to be embraced by almost all practitioners of optical science. During the eighteenth century the paradigm for this field was provided by Newton's Opticks (pp. 11-12).

These insights into the working of science simply underscore the determining nature of theories, models or paradigms. Scientific models embody theories. And theories are types of "conceptual gestalts" which operate in a manner similar to a world view: two individuals may observe identical phenomena but may "see" different things because they reference them to different concept labels. Constructs are the constructed and consciously defined concepts interwoven into all scientific theories. They identify the nodes or junctions in a theoretical network of relationships.

Constructs play the indispensable role of tying theories to observables, although in practice they may be given no more than "partial and shifting anchorage in concreta...(being)...terms which, though not observational either directly or indirectly, may be applied and even defined on the basis of observables"

(Kaplan, 1964, p. 55-56). Various constructs such as ability or aptitude are inferred from the results of tests, just as velocity is inferred from the distance traveled in a certain time. And inferring the existence of a construct implies the use of theory.

Theories and constructs, therefore, live in a symbiotic relationship: theories making contact with observable reality through the vehicle of constructs; and constructs claiming existence from theoretical inferences. Single constructs or measures of that construct do not operate in isolation but within either an explanatory or perceptual system, referred to as a nomological network, model, or paradigm (Cronbach and Meehl, 1967). In turn these conceptual structures are grounded in some form of theory whether well articulated or not. Constructs place theories and observables on a continuum; in science there is never a purely empirical observation, as "what is observed" is determined by a theoretical stance. For example, the biological concept of epigenesis cannot make sense outside a carefully delineated theory of ontogenetic evolution explicating the passage from genotype to phenotype. Facts are dumb until they are interpreted; but they can be made to speak so as to support a variety of theoretical stances: an IQ score may be interpreted as reflecting innate ability and potential, or the mere results of enviromental factors such as schooling.

Kaplan (1964) refers to one aspect of this reciprocal relationship between constructs and theory as the "paradox of

conceptualization"; "proper concepts are needed to formulate a good theory, but we need a good theory to arrive at the proper concepts" (p. 53) . This is precisely the paradox of construct validation: reading testing and reading theory were in just such a circular enterprise in the 1940's and 1950's. Scores on norm-referenced reading tests--processing little or no construct validation--were factor-analysed in the vain hope of arriving at some sort of an explanation of reading (see Davis, 1944; Holmes, 1953; and Singer, 1970). To have come to fruition these investigations should have explored some sort of a theoretical explanation of reading. Theories are not definitive statements, but rather temporary heuristic devices under constant revision. However, in the meantime "They allow you to fit together pieces of information about a behavior complex (e.g., reading comprehension) in an attempt to better understand and illustrate the interrelationships" (Lehner, 1979, p. 319). This enables purposeful research to proceed in the face of the "paradox".

The paradox is resolved by a process of successive approximations: "the better our concepts, the better the theory we can formulate with them, and in turn the better the concepts available for the next improved theory" (Kaplan, 1964, p. 53). Each step along the way facilitates a closer match between constructs, theory and reality; and is articulated through more and more exact definitions of constructs. This series of approximations is dubbed as a"bootstrapping operation"

by Cronbach and Meehl (1967). It operates as a constant upward spiral of knowledge directed by the scientific method.

To conclude, constructs are tentative summaries of reality. A construct label on its own is of little value without its accompanying theoretical framework; therefore, the construct validation of any test is only as good as its underlying theory. This implies the relativity of construct validity. A construct may be adequately measured in terms of the theory, but the latter may fail to explain the facts in a satisfactory manner.

Because the logic of construct validity is isomorphic with the logic of science, the process of construct validation is no more or no less than "doing science" -- it is "integrated with hypotheses testing and with all of the philosophical and empirical means by which scientific theories are evaluated" (Messick, 1975, p. 956). Some of the scientific procedures involved in establishing construct validity will be discussed in the following section.

Ways of Establishing Construct Validity

Amassing Evidence

Phrases such as the "gradual accumulation of information from a variety of sources" are a constant refrain in the discussion of the type of evidence necessary to throw light on the nature of the trait being measured and the adequacy of the measure. (Anastasai, 1968, p. 115). Loevinger (1967) appeals for two types of evidence: the first to demonstrate that the test measures

something systematic and the second to lend support to a particular interpretation of a test score. For his part, Messick (1975) appeals for convergent and discriminant evidence (see Cambell and Fiske, 1967). This technique is recommended as each construct measurement represents a trait and a method unit, and often different measures of a trait measure different entities (see Kendall et al., 1980; and Ward et al., 1980). As Cronbach (1971) emphasizes: "One validates not a test, but an <u>interpretation of data arising from a specified procedure</u>" (p. 447, emphasis in the text). Thus, fresh evidence must be furnished to justify each change in procedure or purpose.

There are very few published studies in the social science literature where construct validity is the central concern of the investigation (Carmines and Zeller, 1979). In recent years, however, with the growth of interest in in-depth explanations of test scores a few paradigm studies attempting to explicate the cognitive processes invloved in taking the test have appeared. For example, Ward et al.(1980) investigated the validity of two forms of a test of formulating hypotheses. They began their investigation by speculating on the cognitive processes, and knowledge and personality factors that might be involved in formulating hypotheses. Since no such theory exists, these authors arrived at the factors which might help an examinee formulate satisfactory hypotheses through "introspection" and "armchair speculation " (p. 14). A sample of university students took the multiple choice and free response versions of the test of

formulating hypotheses plus a battery of existing tests representing underlying cognitive, knowledge and personality factors. The correlations between the two test forms were low, hence they could not be considered alternate forms. These low correlations were explained by factor-analysing the results of the tests representing the cognitive, knowledge and personality factors.

Considerations of purpose were a major concern in the conclusions drawn from this study: if the purpose of the test is to select people for traditional academic pursuits, the multiple-choice version of the test has no particular value in supplementing selection instruments like the Graduate Record Examination. "The free-response test, on the other hand, merits further consideration. The production of ideas depends heavily on abilities other than those which determine performance when the subject has only to evaluate alternatives which are presented for choice" (Ward et al., 1980, p. 271). But measuring these abilities may be important for certain purposes: "Real problems, in science and in life, rarely present themselves in multiple-choice form. Second, there is a perennial interest among educators in teaching and testing for creativity, along with a large body of theory and research suggesting that divergent production abilities have some relevance to the subject" (P. 271). As a consequence, identical scores on these two test forms can mean very different things.

Since "questions of test bias or the adequacy of measure-

ment" and "test fairness or the appropriateness of use" (Messick, 1980, p. 1, 012) have become live issues, increasing interest has focused on how different populations interact with a particular test. In a study of the construct validity of functional reading tests, Kirsch and Guthrie (1980) examined the task "in terms of the processing demands which resulted from structural variables of the task" (p. 90) but then went on to explain how people from different educational backgrounds extracted information from the various formats. As a result of their study these authors remind us that reading tests do not have generally accepted meanings.

Although eschewing overt process explanations and direct mention of construct validity another group of researchers investigated systematic measurement error when tests were administered to ethnic groups other than the cultures for which they were standardized (see Cronbach, 1975, for a discussion of this work). Eells (1951) carried out the first major experiment in "culture fair" intelligence testing by changing the content of items. Prompted by the work of Labov (1970), other researchers have concentrated their efforts on rewriting standardized tests in non-standardized dialects. Labov et al.(1968) for example, found that lower class black youths performed at or above "normal" when tested in their own dialect. As a result of these studies it became evident that many tests were not measuring what they purported to measure but rather such spurious variables as knowledge of standard English or the possession of a middle class white

background.

None of these authors investigating construct validity or "what tests measure", were able to formulate definite research hypotheses since at present none of these areas are explained by well-developed theory. No definite predicitons can be made in the early stages of construct development, so various exploratory studies must be undertaken: investigations of how different groups perform on tests, studies of correlations among items within tests, process observations of the students as they do the tests, etc. (Allen and Yen, 1979; and Cronbach and Meehl, 1967). What kind of evidence, then, is deemed acceptable as grist for the mill of construct validity depends on the stage of development of the construct and its theoretical framework. The development of any scientific concept usually proceeds from vague intuitive ideas to imperfect measurement and finally to standardization. For instance, in the pioneering days of the Standford-Binet teachers' judgements were used to choose the test items; nowadays a score on the Standford-Binet Intelligence Scale is awarded more credibility than a teacher's judgement. An instrument may fortify its construct validity simply through repeated use and anecdotal evidence testifying to its trustworthiness.

However, a more credible form of construct validity is established as the nomological network is advanced through research findings and with each successive advance more and more specific hypotheses can be derived and tested. Validity is then judged on the compatibility of the predictions and the findings. Kirsch

and Guthrie (1980) maintain that the strength of a test's "construct validity is directly proportional to the accuracy with which the cognitive requirements for the test predict performance" (p. 91). However, "Any testable prediction can be made to support constuct validity, including predictions of content and criterion-related validity" (Allen and Yen, 1979, p. 108). When the investigations confirm the predictions, the progress of construct validation is relatively unproblematic; however, if the opposite is the case, conclusions are not as clear-cut.

In the event of negative evidence, the researcher "may conclude that the theory is incorrect, or he may equally well conclude that the operationally defined variables are, after all, not really measures of the constitutively defined variables, but are only presumed to be related to them, and hence, that the theory may still be correct and the experiments simply inappropriate" (Torgerson, 1958, p. 7). Thus, incompatible evidence presents a dilemma which canbe resolved only by informed human judgement. Possible decisions range from plans to abandon that particular measurement of the construct, to conclusions that the statistical technique was faulty (Cronbach and Meehl, 1967). The route chosen depends on rational factors such as the import of the negative evidence or the strength of the competing hypotheses but rests also on irrational factors such as attachment to a "pet theory" (Chamberlain, 1964). Although a construct may brave the onslought of conflicting evidence, it may never be proven to

exist in an absolute sense. This follows from Popper's (1959) observation that theories are falsifiable not provable; but also from the previous discussion about the nature of science and scientific knowledge.

Because of the nature of science, construct validation cannot be a once oft affair; rather, it is an ongoing process. "(C)onstruct validity is not established by confirming a single prediction on different occasions or confirming many predictions in a single study. Instead, construct validation requires a pattern of consistent findings involving different researchers using different theoretical structures across a number of different studies" (Carmines and Zeller, 1979, p. 23). Evidence may be sought by specially designed validational studies, but more fruitfully from advances in research in the basic disciplines. Generally this has not been the case. Enright and Lapsley (1980), for instance, demonstrate that the psychometric and the social cognitive stage theories of role taking have developed relatively independently.

A symbiotic relationship should exist between psychometrics and the basic disciplines within pscyhology. The process of test validation should be concerned with the discovery and validation of substantive psychological knowledge; this is similar to one of the main points of Cronbach and Meehl's (1967), and Loevinger's (1967) papers. Carroll's (1978) characterization of psychometrics as a technology deprives the testing enterprise of its true potential.

To conclude, construct validity may be viewed as inference from evidence, with different inferences requiring different evidence. Thus, by "focusing on the nature of the evidence in relation to the nature of the inferences drawn from the test scores,...validity (may be viewed) as a general imperative in measurement. Validity is the overall degree of justification for test interpretation and use" (Messick, 1980, p. 1, 012). Furthermore, because scores are a function of subjects' responses, evidence should help explain the processes underlying these responses as well as correlated factors which may affect the functioning of these processes such as age, ethnicity or level of test-taking sophistication. These concerns, however, are voiced primarily as post facto problems of test validation, not as caveats which should be incorporated into test construction. In existing models of test construction, construct validity is not generally an issue until the test is already constructed.

Test Construction

The omission of considerations of construct validity from test construction appears to be widespread, because, with a few exceptions, theoretical discussions of construct validity are concerned only with post facto validation of tests (e.g., Cronbach, 1971; and Messick, 1975, 1980 and 1981). Loevinger's (1957/1967) monograph is among the exceptions, as her theory of construct validity is nothing more than a complete theory of

test construction presented in rough outline. Test items must be generated from the most satisfactory theoretical understanding of the construct, be it anxiety, self-concept, or reading comprehension -- because in due course how a construct is measured will be used to redefine theoretical accounts.

This circularity is nothing more than the paradox of conceptualization (Kaplan, 1964); a description of something is based on a prior conceptualization, but in turn this theoretical construct cannot be verified without the empirical evidence which is gathered with a conceptually biased apparatus. As the nature of measurement and theoretical explanation are so inextricably woven, every aspect of test design and construction should be automatically considered a part of construct validation if constructs are being assessed.

Present or past models of test construction have either explicitly or inadvertently delayed mention of concerns of construct validity until the instrument is already constructed. In classical test theory the items are constructed from a table of specifications and revised through an item analysis; reliability and validity are then established, typically through some form of correlation study (see Allen and Yen, 1979, pp. 56-147). Recently classical theory has been replaced by the latent trait theories for the construction and analysis of large scale achievement tests. However, because of the normal ogive, unidimensionality, and independence assumptions limits are set on the kinds of items that may be included in a test. Instead of discovering

inherent relationships, a Rasch or a Logistic analysis imposes an alien structure on the relationship of ability and achievement. Thus, concerns of construct validity are automatically ignored in test construction except in rare circumstances where a trait fits all the model's assumptions. As a consequence, these tests are not suited "to serve as instruments of theory" (Loevinger, 1967, p. 99).

In contrast to the classical and latent trait models introduced above, much of the published work on criterion- or domain-referenced tests is presented in the form of practical guides for the construction of tests referenced to a clearly specified domain of content, behavior, or objectives (e.g., Berk, 1980; Hively, 1974; and Popham, 1976). "Most of the theoretical and empirical work in Domain-Referenced Testing is devoted to making the concrete domain more and more representative of the essential skills in the original universe" (Hively, 1974, p. 10) Baker (1974) defines domain as "a sub-set of knowledge, skills, understandings or attitudes where the essential attributes of the content which the student is expected to acquire and the behavior through which he or she is expected to demonstrate such acquisition are carefully described" (p. 10).

The focal point of this work is on content relevance resulting in a deemphasis on process or interactional questions. Most authors writing about domain- or criterion-referenced tests claim, directly (e.g., Hambleton et al., 1978) or indirectly (e.g., Wardrop et al., 1978), that the basic validity requirements

are issues of content referencing and generalizability of scores.

But domain or criterion referencing is not a sufficient condition for construct validation because regardless of the sophistication of techniques of domain description and referencing, the main emphasis is on item content. Kirsch and Guthrie (1980) remind us: "The construct validity of a test score is not established by a description of the structural features or underlying processes on which items may vary" (p. 91). Furthermore, "interpretations claiming content validity...must be carefully restricted to task language; to the extent that attribute or process language is used, construct validity is required" (Messick, 1975, p. 959).

Although not a suffient condition, some form of domain referencing is a necessary condition for construct validation. However, a fundamental problem arises for tests referenced to domains in the traditional fashion when it is realized that it is not possible to demarcate a finite domain. Loevinger (1965) cogently argues that it is impossible to delimit unequivocally a domain of content or behavior, as "there are unknown and uncontrollable aspects of the test situation, (also, because a particular domain reference changes for each test score as each individual with his unique background interacts with an item), but one works always towards optimization" (p. 148) when constructing tests. Hence, tentative domains must be explicated in interactional terms; it is differences in <u>how</u> examinees process the items that cause differences in test scores. Since "how examinees process the items" is the

construct that is being measured, it logically follows that in the case of test construction domains and constructs are one and the same; domains like constructs are only tentative summaries of an aspect of perceptual reality.

When from the initial stages of item construction, a test is referenced to a tentative domain of processing, the notion of examinee test interaction is of foremost concern. As a direct consequence, construct validity becomes the guiding force in all phases of item creation and revision, but conversely, a measure of construct validity may be claimed for a test because of how it has been devised. However, this is not the case for criterion- or domain-referenced tests because of inadequate concepts of test domains. To conclude, instruments referenced to tentative domains may lay a direct claim to construct validity as these measures are construct-referenced.

Models of Test Construction

The Model of Wardrop et al.

In spite of the presence of models or plans for the construction of various types of tests, there is an absence of comprehensive models of test construction which would offer a general perspective on the field of psychological measurement, and also attempt to relate the end purpose of a particular measure to the appropriate procedures, thus presenting a rationale for these procedural choices. The model of test construction devised by Wardrop et al. (1978) is an attempt to fill such a gap for

reading tests. These authors present a holistic framework of the most important dimensions that characterize these tests. At present, this model is far from ideal. However, it provides a base or a framework within which to discuss the procedures which must be considered when constructing a measurement instrument.

A Framework for Constructing and Analsying Reading Tests

The model of test construction of Wardrop et al.(1978) attempts to map out and interrelate the concepts and tools of psychometricians as related to reading. According to the authors, this model is intended both to describe and evaluate reading tests, and to prescribe the appropriate steps in constructing a test for a specific purpose. The model incorporates five important dimensions which should be characteristic of any test following the guidelines set forth in the _Standards for Educational and Psychological Tests_ (American Psychological Association, 1974). These five dimensions are: uses of tests, item generation, item revision, assessment of precision, and validation procedures. Figure 2, reproduced from Wardrop et al,(1978) graphically portrays these five dimensions as continua running from the position of the pure form norm-referenced test on the left to the pure form domain-referenced test on the right hand end-point. The dimensions will be introduced plus an explanation of Wardrop et al's concept of "consistent inference system" for tests.

Uses of Tests. As as rule, the uses of reading tests

USES OF TESTS

Differentiation					Measurement
maximizing job performance	fairness in job allocation	minimizing effort and disappointment in training	certifying competence	diagnosing strength and weakness	tracking progress

ITEM GENERATION

Descriptive categories			Generative rules
table of specifications	list (catalog) of objectives without theory	ordered list of objectives	theoretical partitioning of specified set

ITEM REVISION

Focus on items		Focus on rules
selecting and fine-tuning items	adjusting objectives	modifying rules, theories for generating or selecting items

ASSESSMENT OF PRECISION

Intersubject		Intrasubject
one time measure, intersubject variability, internal consistency	generalizability theory	repeated individual measures, time series, function fitting

VALIDATION

External				Internal
correlation with external related criteria and tests	content	face	construct based on external and internal criteria	structural

Figure 2. Framework for Test Construction

can be classified into either of two major clusters. The first cluster centers on relative selection or differentiation between individuals; the second focuses on absolute assessment and measures individual competence. Therefore, the first dimension, test uses, ranges from tests whose purpose is to differentiate between individuals depicted on the extreme left of the continuum, to tests whose concern is with intra-individual measurement, on the right-hand extremity.

The prototypic norm-referenced test is situated at the left, as the main purpose of these tests is to differentiate between the performance of various individuals. A more specific use is to place people in discrete categories: "Differentiation typically is the basis for assignment of examinees to categories (hired-rejected; admitted-denied; gifted-normal), that is, for making selection decisions when access is limited" (Wardrop et al., 1978, p. 4). Usually these tests are administered to large numbers of children or adults to perform such functions as fairly selecting between applicants for educational and/or vocational positions, thereby minimizing undue effort and disappointment in school learning and maximizing job performance.

In contrast, the prototypic domain-referenced test is situated at the right-hand endpoint. All the right-hand side of the scale of uses is concerned with tests which emphasize measurement, that is, estimate the quantity and/or quality of a characteristic within an individual. Criterion- or domain-referenced tests are designed to measure how well an individual

performs with reference to a domain of behavior. Typically, these tests lead "to decisions about each individual in a situation where access may be, at least in principle, open to everyone (pass-fail; certified-not certified; given new instructional material-recycled through previous material; assigned activity X-assigned activity Y). Here the purpose is not to choose among many but to provide guidance to each individual" (Wardrop, 1978, p. 4).

Item Generation. The item generation dimension runs between items which are constructed from "descriptive categories" and items derived from "generative rules". As in the first dimension, norm-referenced tests are situated on the left-most extreme. The items used in this type of test are generally a sample chosen from a table of specifications which is based on a descriptive analysis of content. In contrast, each item in a test of measurement is derived from a set of generative rules, which in turn are derived from a theory of reading or at least from an ordered set of objectives. Once again the right-most pole of this dimension is defined by the typical domain-referenced test.

Item Revision. This dimension locates the ways test items are revised, after piloting, on a scale ranging from techniques which center on adjusting and fine tuning individual items to techniques which emphasize making changes to the generative rules which directed the item generation. Thus in line with the

test placement on the two previous dimensions, norm-referenced tests anchor the left end-point and domain-reference the rightmost one. If in the field testing of a norm-referenced test it is discovered that items fail to differentiate between individuals or are too easy or too difficult, the problematic items are then replaced or adjusted. If however a domain-referenced test fails to measure an individual's performance, the rules used in item generation are first modified, then the individual items are themselves modified via these rule changes.

Assessment of Precision. Assessment of precision refers to how measures of reliability are established. The accuracy and reliability with which a test measures what it is designed to measure is an essential dimension for judging the quality of that test. The scale depicting how tests vary in terms of how reliability has been established has at its two extremities: (a) procedures based on intersubject variability which accounts for the norm-referenced tests, and (b) procedures based on intra-subject consistency which are used in domain-referenced tests.

According to Wardrop et al, all standardized, norm-referenced achievement tests and most other tests rely on techniques based on inter-subject variablity to obtain estimates of reliabilities. "The exact procedures differ somewhat from test to test, but essentially the goal is to estimate the reliability with which the test will consistently differentiate among people" (p. 17). Included in these estimates of precision and consistency are test/

retest, alternate forms and split-half reliabilities (see Allen and Yen, 1979, pp. 76-83).

For tests intended to provide measurement data, these classical notions of reliability become inadequate. A domain-referenced test is judged reliable if it is able to sensitively assess an individual's performance over time. Wardrop et al. refer to this as "repeatability". Repeated testing with a domain-referenced test would yield a smooth growth curve if performance has improved, and a horizontal line if the performance is stable over time. Appropriate techniques for this kind of an assessment include time series analysis and function fitting. Implicit in the concept of reliablity is the existence of a large number of test forms, all measuring the same behavioral domain.

<u>Validation Procedures</u>. The final dimension of the model represents validational procedures and ranges between the two end-points of external and internal procedures. For tests used to differentiate, considerations of external validity such as content and predictive validity are paramount. And for tests of measurement, internal validity considerations, such as construct and structural, are of primary importance. Therefore, norm-referenced and domain-referenced tests can be placed at the left-hand and right-hand end-points. In all, five types of validity have been identified on this dimension of the model.

The concept of predictive validity presented in this model is best represented by the notion of concurrent validity.

A concurrent validity coefficient is obtained by administering the new test at the same time as some well-established test, and correlating the results. The second point on the scale is content validity. "An educational achievement test may be said to have content validity to the extent that it measures those goals which are the goals of instruction in that area" (Wardrop et al., 1978, p. 20). At the third point on the scale is face validity; it is considered by the authors as "a concept of limited value" and is sometimes called "armchair validity".

The two stages on the right of the scale--construct and structural validities--are, according to the authors, "closely related" forms of validity. The definition of construct validity is the same as that developed previously in this chapter. Structural validity refers to Loevinger's (1967) term, and "applies to tests that have clearly defined procedures for creating items. If the structure of a test is derived from some model of the behavior to be assessed, the structure itself should reflect those features of a task presumed to affect the difficulty of the task ... In such a situation it is possible to infer from the test structure what specific outcomes should occur in any research in which the variables used to create the tests are manipulated" (Wardrop et al., 1978, pp. 21-22).

This distinction between the general concept of construct validity and the notion of structural validity is puzzling. Possible reasons for this distinction will be developed in the section devoted to describing some limitations of this model.

Inference System for Tests Derived from the Model. The model presented above is intended by its authors to have both a descriptive and a prescriptive function. Existing tests can be plotted on the various dimensions, and this analysis of test characteristics can lead to the judgment that there are "good" and "not-so-good" profiles of test characteristics.

> In particular, a straight line running down either the left side or the right side of the set of scales in the descriptive model represents a consistent inference system. That is, item generation, test revision, assessment of precision and validation procedures are all designed and carried out in such a way that they support the use for which the test was intended. These four characteristics are logically dependent on the intended use of the test. Tests or testing systems which are represented by lines running in a zigzag pattern across the scales suggest inconsistencies in the systems of inference they are using. Most existing tests show a zigzag pattern to some extent (Wardrop et al., 1979, p. 37).

Therefore, a "consistent inference system" consists of a set of test characteristics that can be plotted directly below each other on each of the five dimensions and thus can be graphically represented by a straight line running down either the left, right, or centre of the model. According to Wardrop et al., this has implications for test construction; once the test use has been identified, the types of techniques and procedures involved in item generation, etc., can be automatically chosen by identifying points that lie directly below on the other four dimensions. In this model the potential use of the test determines the underlying model of test construction.

Two rather straightforward restrictions can be derived

from the above interpretation of the testing system: (1) it is not possible to change the purpose of tests from one end of the scale to the other -- for example, it is not permissible to use a norm-referenced test to diagnose reading difficulties nor a domain-referenced test to select among students; (2) it is possible to have a test that serves both purposes only when both supporting inference systems are incorporated.

Limitations of the Wardrop et al.Model

The model of test construction of Wardrop et al.(1978) has been introduced in the present work because it appears to be the best available model of its kind; but also more specifically because the model which will be employed to direct the construction of the instrument to measure reading processes is a revised version of this model. Because of some limitations, which will be pointed out, the inferential system posited by the authors will not be strictly adhered to. Hence, the model's descriptive and categorical qualities and not its dynamics are appreciated. The second, more specific, criticism refers to aspects of the graphic model and concerns validity.

Wardrop et al.structure their model by using ideal norm- and domain-referenced tests to anchor the end-points of the five dimensions of test characteristics. For these authors, there appears to be an associative meaning inherent in these labels. The labels are used as if they ideally and inherently possessed the five characteristices at opposing ends of the scale. Yet

elsewhere, Wardrop et al.(1978) see the intended use of the test, not the label, as the determining factor when it comes to profiling a test on each of the five dimensions.

This ambiguity concerning test purpose and test labels results in a test being situated on the first dimension of the model, test uses, according to purpose not label; then, however, its position on the other characteristics automatically predetermined. Decisions should be made to determine the criteria which will affect the choice of each of the test's characteristics. For each decision the determining factor should be considerations of whether that particular characteristic suits the test's purpose. Therefore, test purpose, and not labels or how tests are named, should determine the criteria which will affect the choice of each characteristic within the model, as some characteristics may possess more flexibility than others. For example, item generation has less flexibility than other dimensions such as reliablity. Wardrop et al.'s points concerning procedures of item generation for norm- and domain-referenced test are valid for the traditional <u>uses</u> of these tests. However, procedures for the estimation of reliablity for a test with two forms (traditionally a norm-referenced test) would appear at the left-hand side of that scale. But this researcher sees no reason why this type of reliablity cannot be applied, as long as the criteria of "two parallel forms" is met, regardless of whether the test is designated as norm-referenced or domain-referenced. Furthermore, there is no good reason preventing a test, whose primary purpose is to differentiate

among students on the basis of achievement, from having construct validity.

As a result of such considerations, the five characteristics may range at various points across the five scales. Thus, the notion of inferential consistency cannot be seen in terms of polarizing straight lines, but in terms of making each placement consistent with the criteria established for each test, and in accordance with its use.

The second major criticism of the model of test construction concerns the role of validation and, in particular, of construct validation. The placement of the validational dimension after concerns of item creation and revision, and reliablity implies that a test must first be constructed before its validity can be considered; this placement may not have resulted from a conscious decision and may be but an artifact of the graphic presentation. Yet, within this dimension structural validity (borrowed from Loevinger, 1967) is positioned to the right of construct validity thereby <u>inadvertently</u> implying that a measure with demonstrated structural relationships with the parent domain is better suited to gauging the quantity and quality of a trait than a more general type of construct validity. For Loevinger (1967) structural validity is only one of the three sub-components of construct validity which focuses on the characteristics of items. Wardrop et al. (1978) probably give more weight to structural validity and item content, being advocates of domain-referenced testing.

Since the model is a model of test construction, the

complaint is not that validity is considered unimportant in test construction; but it is rather a question of its relative importance and integration within the complete model. Although theoretical concerns are included in the item creation and generation phases of the Wardrop et al. model, these design phases are not depicted as an integral part of construct validation; and so are not set forth as important facets of the validational evidence. In addition, the other aspects of validity such as concurrent and predictive validities rather than being independent of, may also be utilized as evidence towards construct validation.

A Model of Construct-Referenced Test Construction

Considerations of construct validity are the focus of the present study; hence, it follows from earlier discussions that this form of validity must direct the construction and interpretation of the measurement instrument right from the initial phases in the conceptualization of a prospective measure. Thus, a construct-referenced model of test construction will be presented which will be capable of incorporating the theory of reading processes into all aspects of test construction. In this way measurement theory and reading theory are conjoined in the pursuit of scientific knowledge. This model is presented in Figure 3 and maintains many of the concepts of the Wardrop et al. (1978) model.

The revised model is suitable for the construction of an instrument intented to measure how an examinee engages in a particular task. Measurement purpose should determine how a test

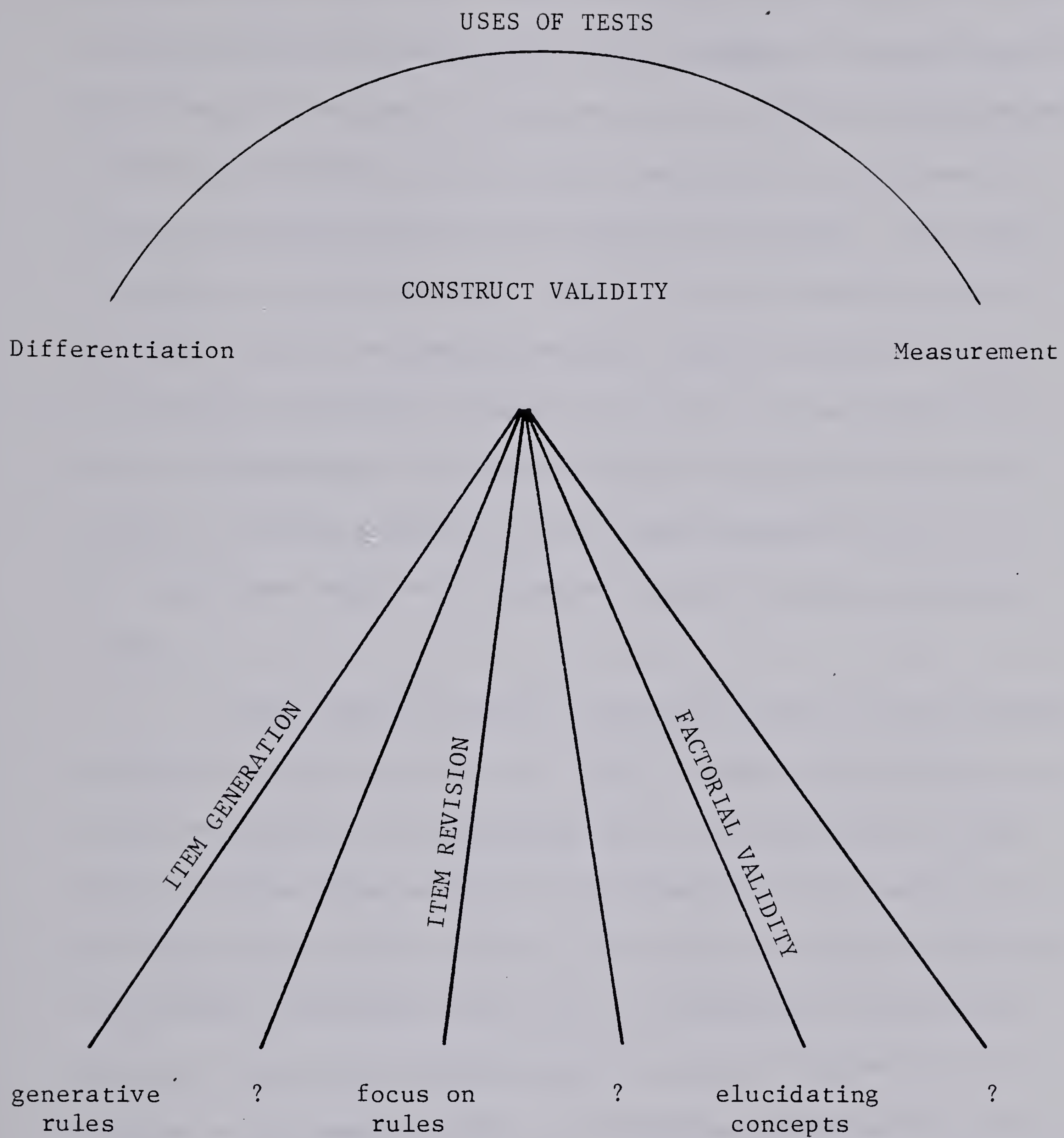

Figure 3. Model of the Construction of Construct-Referenced Tests

is validated, so the "test uses" dimension is the only one of the five major model characteristics which remains in the same position. This dimension may now be conceived of as an umbrella under which the other major decision will be made, namely, how the assessment of reading processes will be constructed and validated. The second dimension is the validational dimension, but in keeping with the purposes of this measurement project, construct validity is the overriding validational concern. Moreover, it was shown above that considerations of construct validity may subsume the other forms of validity for this or any other assessment, as well as the procedures employed in constructing and revising assessment items.

The construct validity dimension consists of an infinite number of possible subordinate nodes because establishing construct validity is an unending process. For the present study, three of these nodes have been identified--item generation, item revision, and factorial validity; these are the specific forms of validational procedures which will be undertaken in this study. However, some nodes are left unnamed and may be filled in future studies -- for example, future investigations might involve research to establish concurrent or predictive validities.

The item generation and revision procedures are those named on the measurement end of those dimensions of the Wardrop et al. model. Items are first derived from a set of generative rules specifying the content and structure and naming the variables which may be varied to increase the processing demands. When the items

are revised the focus is on changing the rules rather than on blind manipulation of individual items. However, changes may be made to some items without rule changes, when piloting shows there has been a failure to implement these rules.

The third sub-ordinate node of the present validational dimension, factorial validity, marks the first known attempt to amass evidence to demonstrate the construct validity of the instrument devised to assess reading processes and its supporting theory.

Reliability or assessment of precision, has not been named as a distinct dimension, as concerns of accuracy and precision are at stake in each stage of test construction and subsequent evidence-gathering studies. And as Loevinger (1967) pointed out, some of the traditional techniques of assessing reliabilty such as split-half or parallel form reliabilities should be considered as a part of construct validitation. In the present model, these studies would then be added as additional subordinate nodes in that dimension.

Conclusion

In this chapter a psychometric rationale was developed for the construction and validation of an instrument designed to measure the construct of reading outlined in the theory of the assessment of reading processes. Construct validity and therefore construct referencing were defined as indispensable routes to the elimination of systematic error in psychological measurement, thus verifying that the constructs being measured have a basis in reality. The

ongoing collection of validational evidence is one of the principal means of elaborating a nomological network for the construct. But in addition, it was argued that if tests are to be considered as instruments of theory development, each phase of test construction must be conceived as a component of the validational evidence.

As a consequence of these arguments, a model of test construction was built to incorporate item generation, item revision and a factorial validity study into the process of construct validation. This innovation is merely what Loevinger (1967) pleads and is demonstrated by Jackson and Messick's (1967) succinct summary of that plea for construct validity:

> Jane Loevinger's classic monograph argues not only that test and theory validation ought to go hand in hand but that tests, if they are to represent instruments of psychological theory, must be developed systematically in terms of that theory. The validation process implies a program of test development in which substantive theory plays a dominant role at every stage of the process, from definition of variables and item writing to evaluation of the tests structural properties and external validity. In this analysis, Loevinger shows little patience with approaches that allow test developers refuge in mechanical, _ad hoc_ procedures like empirical item selection (p. 42, emphasis in the text).

CHAPTER V

PROCEDURES: ITEM GENERATION AND REVISION

Introduction

The purpose of this chapter is to describe the procedures which have been followed in constructing an assessment instrument suitable for measuring reading processes. The ultimate goal of this test construction enterprise is aimed at the collection of data which will serve as a further increment of evidence in the ongoing campaign of construct validation of the theory of assessment of reading processes. However, since the construction of the instrument is based on the theory of the assessment of reading processes, the procedures used to derive it from the reading theory are an integral part of the construct validation of both theory and instrument. This instrument should be capable of measuring eight processes which are assumed to take place during reading. In consequence, the proposed assessment will henceforth be referred to as the Assessment of Reading Processes (ARP).

The revised model of test construction introduced in the previous chapter will be used to structure the construction of the ARP (see Figure 3). Specifically, the two dimensions of the construct-referenced model, test uses and construct validation (with two of its components: item generation and revision), will be employed to organize the presentation of the procedural stages. The main body of this construction will therefore be included in the validation category; as a consequence the theory of reading

processes will be brought to bear on all design-focused decisions, thus forging a firm bond between psychometrics and reading theory.

Test Uses

The main purpose of the ARP is to measure eight reading processes; because these are hypothetical constructs and are thus unobservable directly, a theoretically based inference system must be incorporated into the measure. These decisions unequivocally designate the ARP as a "construct-referenced" measure (Messick, 1975, p. 95).

Construct Validity

In the previous chapter various procedures for establishing the validity of a test were discussed, including both test construction and the collection of evidence. For this chapter, three phases of construct validity were identified for the ARP. Two of these aspects, item generation and item revision, which are directly concerned with the present phase of test construction, will be discussed in this chapter; since the third aspect, factorial validity, forms a distinct study it will be discussed separately in the next chapter.

Item Generation

The construct-referenced approach to assessment results in a set of items that are firmly based on theoretical propositions explaining reading processing; this relationship is achieved through a set of generative rules or criteria.

Four distinct steps in the generation of domain-referenced items have been identified by Anderson et al. (1978). Although these authors interpret domain as a domain of content or items, these stages can be adapted for use in a theoretical or construct-referenced domain. These four procedural stages are: (1) identify a plausible candidate domain, (2) identify prototypic items from that domain, (3) specify relevant variables thought to affect the difficulty of responding correctly to those items, and (4) create sample items.

Domain Identification. The theory of reading assessment postulates a construct of reading as a set of mental operations or processes. These include processes that occur in the word identification and comprehension phases of reading. In other words, these processes are involved in the interaction between the various kinds of textual information and the reader's knowledge of the world, language and the reading task, and enable the reconstruction of the author's message.

Thus, the domain of this assessment is focused on these processes which may be inferred from various forms of reading product. As these processes will be inferred from reading product, they may have occurred at the reception of the message or at the creation of the product, or at both stages. A rationale for the existence of these processes is based on the theory of the assessment of reading processes which was introduced in Chapter 3 . These processes are: attending, analysing, associating, sequencing, predicting, monitoring, synthesizing

and inferencing; and the measures or variables which will be used to infer these are called processing indicators.

Since these interactional processes are affected by the degree of match between the knowledge and language of the reader and the content of text, the domain was delimited to involve only the interaction between the reader and text when the reader is reasonably familiar with the structure, topic, syntax, and vocabulary of the text. More specifically, the domain was defined as the processes which occur when a Canadian elementary school child reads a passage following a familiar structure on the topic of sports or animals.

Prototypic Items. Items were defined to "include directions for (1) presenting the stimuli, (2) recording the response, and (3) deciding whether or not the response is appropriate" (Hively, 1974, p. 10). As an informal reading inventory is the primary manner of assessment, it was chosen as a prototypic item. This includes a series of graded narrative passages ranging from grade one to six reading levels, plus directions for presenting the passages and scoring the oral reading, recall protocols. The particulars of writing the passages and the scoring system employed will be discussed later. However, the two sets of directions are given below: (1) for when the child is asked to read the passage, and (2) for eliciting the recall after a five minute interval:

1. I want you to read this passage about... (insert topic). Later I will ask you to tell me the

story. Read it out loud/silently. You may begin now.

2. Now, could you tell me the story that you read.

Variable Affecting the Difficulty of Stimulus Items.
As a result of an analysis of the domain of the theory of reading assessment, four variables were considered to contribute to the difficulty of the reading passages:

1. passage topic
2. vocabulary including within passage word frequency, familiarity of words and ease of decoding
3. passage structure including both the story structure and semantic organization
4. syntactic complexity including the average number of words per t-unit and difficulty level as established by a readability formula.

Criteria or rules for the construction of the passages were derived from the four variables listed above. These generative rules are given below:

1. Passages for each grade will be written on two topics: sports and animals. Topics will increase in difficulty by progressing from the more universal sports to the more specialized, and from the more common to to the more exotic animals.
2. Within passage word frequency will be controlled by ensuring that certain key content words appear across the entire passage.

3. The choice of lexical items will range from the very familiar, concrete and frequent, to less familiar abstract words. Choice of words will be based on the judgement of the test constructor, plus a scaling of the vocabulary in terms of the word frequency index which is based on the words appearing in writing done for elementary school (Carroll et al., 1971).
4. Each passage will contain some "difficult" words. Because one of the modes of analysis of processing is a miscue analysis, some unusual words and some words with irregular spelling will be inserted in each passage, e.g., "bad cough" instead of "cold" in a second grade-passage.
5. Passages will be constructed according to the rules of story grammar (Mandler and Johnson, 1977). Passages will increase in difficulty from grades one to six by adding an extra episode plus character every two grades; also grades five and six passages will contain some "Then" connected episodes.
6. The number of semantic propositions per passage will be noted. For each grade, passage difficulty will be increased by progressively adding to the total number of semantic propositions.
7. The number of words per passage will be counted. The number of words per passage will increase across

the six grade levels.

8. The average number of words per t-unit will be calculated. As the passages increase in grade level, the average number of words per t-unit will increase.

9. A level of relative difficulty will be established using the <u>Fog Index</u>. This will give a rough indication of the difficulty levels of the various passages.

10. A title will be added to each passage. But unlike the other variables it is not intended to increase the reading difficulty across the passages. Therefore, a neutral title simply stating the topic will be chosen for all passages regardless of level. This will provide the reader with a prior set or framework in which to comprehend the passage (see Nicholson, 1977).

<u>Creating Sample Items</u>. This section will outline the procedures that were followed in the creation of the <u>initial</u> set of sample items for the ARP. Many aspects of these items remain unchanged, but the present set of passages and scoring systems has been revised as a consequence of piloting. The revised versions of the passages and scoring systems are included in the forthcoming discussion on item revision. This order of presentation has been chosen in order to highlight the sequence followed in designing the ARP.

This stage of item construction saw the creation of passages for two forms of the IRI and of a scoring system to enable processes to be inferred as a consequence of reading the passages.

Stimulus passages: Twelve passages were written--six on sports topics and six on animal topics. These passages were written as parallel forms--two at each grade level--according to the criteria specified above. Comparison tables were then constructed to check the parallelism of the stories in terms of vocabulary and structure. Appropriate changes were made when necessary. An example of a comparison table for the structure of the two Level-1 stories is included in Table 1, and examples of two pre-pilot stories are given in Table 2.

Scoring system: The goal of the scoring system is to enable the target processes to be inferred from two samples of assessment data: oral reading and unaided recall of passages. Therefore, the system which has been devised may be divided into two parts: (1) inferring processes from oral reading miscues which will be referred to as the miscue analysis, and (2) inferring processes from recall protocols. This will be referred to as recall analysis.

Miscue analysis: As the student reads orally his errors are later copied on to the miscue scoring sheet (see Figure 4). This method of scoring oral reading miscues is a modified version of the Goodman and Burke (1972) miscue analysis. However, these authors did not introduce a system for inferring specific processes

TABLE 1 *

	CHARACTERS	EPISODIC STRUCTURE	SETTING	VIOLATION OF STORY GRAMMAR RULES
The Cat (Level 1)	Two characters acting as one plus the cat.	Basically only one episode but within that 2 characters, child and the cat perform intentional actions.	Actors introduced. No situational information given.	No violations.
The Ball (Level 1)	Two characters acting as one. Some separability of characters required but easy to predict who carries out what action.	One episode, consisting of: beginning attempt outcome ending.	Actors are introduced plus some situational context.	No violations.

* Example of comparison table for the two stories from Level 1.

TABLE 2

Example of two pre-pilot stories.

FORM A THE BALL LEVEL 1

Dan and Kim sat on the grass. It was a nice sunny day but they were sad. They had lost their new red ball. But they wanted to find it. So Kim looked among the flowers and Dan looked in the trees. Dan looked up high. He saw something in a tree. It was the ball. Kim helped him climb up in the tree. He got the ball and then they were very happy. They had a good ball game and lots of fun.

FORM A SWIMMING LEVEL 1

Don is learning to swim. Every Sunday, his sister Pat takes him to the pool. Nine other children are taking lessons, too. They are are all the same age as Don. Last Sunday, Don arrived early. He got changed and waited near the pool. He saw a bright shiny coin under the water. When nobody was looking Don climbed in the pool. He tried to dive to the bottom. Just then the swimming instructor came and saw him. She scolded him because no children were allowed in the pool on their own. Poor Don did not get the coin. But he learned a good lesson. But he was glad because his sister was not there to see him being scolded.

			Graphic					Oral Language				
			Graphic					Syntactic Accept.		Semantic Acceptability		
	Error	Corrected	H	A	L	Unit	Position	Sentence	Partial	Passage	Sentence	Partial
saved	save	+	+			W.	I. M.		+	+	+	+
		Hypothesis testing/monitoring	Attending and symbol-sound associating			Word synthesizing	Analysing and sequencing	Predicting		Synthesizing	Predicting	
1	2	3	4	5	6	7	8	9	10	11	12	13

Figure 4. Miscue Analysis Sheet

from their analysis. Therefore, the present scoring system has been extended to allow such processes to be inferred from the scores.

To facilitate the explanation of the miscue analysis constructed for the ARP, the various columns of the miscue scoring sheet (see Figure 4) have been numbered from 1-13. Columns 3-7 contain information on how graphophonic information is processed, and columns 2 and 8-12 contain information on the use the reader makes of his knowledge of syntax and semantics. The reading processes which may be inferred from the different error scores are indicated directly on the sheet, and will be detailed further below. This description is necessarily brief as this version of the scoring has since been modified (see section on item revision for revised version).

1. The first column simply contains a record of the text words which have been miscued.

2. The second column records the oral reading error.

3. Information about the reader's monitoring behavior is inferred from whether the error is self-corrected or not.

4, 5, 6. The fourth, fifth, and sixth columns indicate the extent to which the reader is attending to graphics and is engaging in sound-symbol association. This can be at various levels: high, average, or low. An error is scored as high (attention to graphics) if two or more parts of the miscue are the same as text;

average if one part is the same; and low if none of the parts are the same. A part is defined as the initial, medial, or final positions of the word.

7. The seventh column yields information on the process of word synthesis, that is, the extent to which the reader is attempting to chunk grapho-phonic units such as phonemes and syllables to produce a meaningful word.

8. The scores in the eighth column allow information about the processes of analysing and sequencing to be inferred.

9, 10. The information in columns nine and ten can help determine if the reader was predicting using syntax at either the sentence or part-sentence levels. A sentence is defined as a t-unit.

11, 12, 13. Columns eleven, twelve, and thirteen yield information about semantic prediction and synthesis.

Recall analysis: During the assessment the child's unaided recall is taped and later transcribed. The protocol is then analyzed according to the system devised by Fagan (in press). The first stage in the analysis is the elimination of mazes and recall conventions. Following this the remaining utterances are divided into clauses. And, finally, each clause is assigned to one or more of the five recall categories.

The recall categories were specifically designed "to

provide a structure to assess the degree of comprehension as indicated by a recall protocol" (Fagan, in press). Table 3 presents the five recall categories plus a synopsis of (a) the type of text information included in each category, (b) the reading and production processes that may be inferred.

The piloting and revision of these sample items will be described below.

Item Revision

Piloting Sample Items. The passages from the first two forms of the ARP were piloted with a sample of twenty-four good and poor readers from grades one to six, primarily to assess the criteria specified for test construction and the appropriateness of the scoring systems. The passage should indicate a progressive increase in difficulty across the various grade levels, and the scoring system should be capable of measuring an individual's processes, that is, allow certain processing strategies to emerge, or indicate changes in processing across the various levels.

A more specific aim concerning the passages was to ascertain if both sets of passages (sports and animals) were generally parallel in terms of the processing they demanded of the reader. Towards this end, both sets were read orally in the pilot; hence, it could be observed if both presented equal levels of difficulty in word identification. Comparative data analyses were carried out on the miscues and recall protocols, although

TABLE 3

Recall Categories

TEXT EXACT Exact recall of text ● Associating--textual information "reproduced"	A
TEXT SPECIFIC Specific references in text ● Associating, sequencing, synthesizing--textual information "transformed"	B
TEXT ENTAILED Summary statements ● Synthesizing (across clauses)--information from within text "reconstructed"	C
TEXT EXPERENTIAL Experential addition to text, e.g., inferences ● Inferencing--related information from outside the test "reconstructed"	D
TEXT ERRONEOUS Information which has been incorrectly processed ● Faulty associating, faulty inferencing, faulty synthesizing	E

no rigorous statistical analyses were undertaken. Since the researcher carried out the pilot assessment, it enabled her to observe how the children were understanding the different levels of passages. For example, it was observed that the children were having difficulty with the names of the characters in the recall, so a slightly modified set of instructions were experimented with and found to work. Following completion of the analyses of data gathered during the pilot, changes were made to the passages, directions and systems of analyses. The revisions and the writing of the additional stories will be presented below.

Story Revisions. Following the pilot study, the children's oral reading errors and recall protocols were analyzed using the two sets of analyses described in the previous section. A chart was then compiled for each story, noting the grade level and reading ability of each child who read it, followed by a list of the child's miscues plus other aspects of the story that seemed to be causing difficulties in recall. An excerpt from one of these charts for Track and Field, level five is included below:

Name	Grade-level	Miscues	Recall problems
Trevor	4; low reader	enthusiastic excells, routine, managed, although, competitions, imagining, unfortunately.	problem with separability of characters; thinks that Frank is practicing for the jumping and has the accident when competing.

The charts for each story were examined to detect patterns of errors, and generally to detect areas within each story that were

too difficult or conversely not posing an adequate challenge to the reader. This detection system was repeated when stories were compared across levels, first to judge the parallelism of the two stories within each grade level, and second to judge if the stories presented gradual progression in levels of difficulty from grades one to six.

The revisions were made for either of these two general reasons: (1) lack of parallelism, or (2) unequal increments of difficulty between passages. The nature of these revisions will be described below:

Vocabulary: When the miscue patterns across various stories were observed it was noted that certain stories persistently produced more miscues than the others at the same level. For example the Track and Field story at level five triggered more oral reading errors than The Zoo; yet for other reasons (to be discussed below) the recall on the former was generally more complete. In Track and Field, for instance, practically every child below grade six had problems with the word "athletics".

Problems arose also because there was too high a concentration of difficult words in one sentence or story part. Once again the most striking example comes from Track and Field: if a child could read past the first two lines without hitting frustration, he read the remainder of the story with relative ease. In the original version the first sentence read: "Frank, Mary and friends are really enthusiastic about athletics." At this stage in the story the children could not make adequate predictions

based on previous context, so they were forced to rely heavily on graphic information. Thus, the overt processing demands became too great.

In addition to certain concentrations of "difficult" words, it was observed that the frequently recurring words within passages were sometimes not distributed across the story. This problem, however, is inextricably tied to questions of structure and referential coherence.

Conceptual load: Although stories were equalized in terms of semantic propositions, it became clear from observing the oral reading behavior and analyzing the recall that some stories presented the reader with a greater conceptual load primarily for two reasons: (1) some concepts were less familiar, so children had problems referencing them to their past experience: e.g., the level six story at present entitled Photographing Wildlife was originally called Photo Safari, which caused few word identification problems, but rather problems of meaning association; and (2) within certain stories there seemed to be a greater range of diverse conceptual referents. As a result, the number of key concepts within each story was noted, and it then became evident that some stories required a greater number of concepts to carry the key ideas, and that some of these concepts referred to a wider range of areas of prior knowledge. In the level-three story, At the Swimming Pool, the action takes place at the swimming pool and consists of two episodes: (1) Peter practices, but Jane watches because she has a cold; and

(2) Jane rescues a little boy who is drowning at the deep end. The resulting conceptual load is broad, and the child has to utilize many areas of prior knowledge to understand the story fully: a schema of illness, a schema of drowning, a schema of rescuing and a schema of the state the drowning boy is in when he is rescued. In contrast the two episodes of the other level-three story, The Dog, take place in the childrens' home, and therefore utilizes a more familiar aspect of semanic memory. This facet of passage difficulty, however, is closely tied to the episodic structure.

Syntax: In a few isolated instances unexpected syntactic structure caused minor problems, and were therefore changed.

Structure: All twenty-four children in the pilot study, regardless of reading ability, were familiar with the story structure; nevertheless some aspects of structure caused comprehension difficulties. In some stories the referential coherence was not obvious to the children: e.g., mid-way in The Budgies (level-four), the budgies were referred to as "chirping friends", and in The Dog (level-three), use of the pronoun "he" precipitated some problems as both Andy and his dog Rex were masculine. As a result of such problems with logical and referential coherence, some action-agent-object confusions were evident in the recalls.

A further difficulty closely related to referential coherence is "separability of characters". At times, readers failed to distinguish among the characters and to connect them

with the actions they performed. In The Zoo (level-five), the three children visit the zoo and choose to report on three different animals.. Many children recalled that all three went to visit all the animals; problems of this nature arose because the actions and the characters were too similiar. Hence, greater diversity was introduced into these stories. Furthermore, some stories required more inferences, although in itself this did not cause major comprehension difficulties but it resulted in the introduction of more new contextually-constrained information in recalls on certain levels, thereby augmenting the representation of certain categories of recalled information. The quantity of information in recalls was affected by two other aspects of story: inclusion of a moral, and quantity of direct speech.

According to the rules of the Mandler and Johnson (1977) story grammar, stories have an underlying or explicit moral. In the piloted draft of the passages, not all stories had a moral; however, when the moral was stated it was usually recalled, but otherwise omitted from recall -- hence causing inequalities in the number of clauses likely to be recalled from stories. In a similar manner, some stories contained more direct speech; and as the direct speech was often recalled verbatim, it meant that some passages facilitated a higher proportion of verbatim information.

To conclude, even though the four aspects of the stories which warranted revisions were presented separately, all are intertwined and all co-determined word identification and coherence

problems; no single aspect occurring in isolation would cause major differences in levels of processing.

As a result of the analysis of the pilot data, it was decided to drop the Fog Readability index as one of the variables which gauged the difficulty of items. The factors which it controlled, such as word length and sentence length, were found not to predict item difficulty. Furthermore, this type of analysis is in contradiction to the concept of test construction as a part of construct validation. As Anderson (1972) cautions, any blind or semi-blind atheoretical manipulation of tests "to control difficulty and discrimination power tortures validity" (p. 164).

Rewriting Stories and Writing Forms C & D. When the dynamics of the problematic aspects of the stories were understood, a revised version of Forms A and B were written as well as two additional Forms C and D; efforts were made to make these as similar as possible to the first two forms. Forms A and C were on sports, and forms B and D were on animal topics. In order to check the implementation of the story writing rules, comparison charts were drawn up explicating the various rules of story grammar which were implemented, such as the number of characters and episodes -- in addition to a listing of the average clause and t-unit length, number of words per passage, within-passage word frequency, and the ten most "difficult" words as verified by an elementary school teacher. Difficulty refers to both phonic regularity and familiarity to children of various ages.

The ten words were rated according to their frequency count in the American Heritage Word Frequency Book (Carroll et al., 1971). However, the final decision to include or to eliminate a word was made by the researcher, taking grade level, orthographic regularity, ease of conceptual association, and frequency count into consideration. Following an investigation of these comparative feature charts, appropriate revisions were made to the four forms. The twenty-four passages were then rewritten, and submitted to the scrutiny of two university professors with backgrounds in reading and many years of experience of the type of reading material to which children are exposed in school.

Both professors made many suggestions as to how the stories might be improved, both from a stylistic and a structural standpoint. Unfortunately, because of constraints of the item writing rules, some of these suggestions could not be implemented. The present version of the passages was written incorporating the majority of these improvements. Examples of passages from the two (A and B) forms of the ARP are included in Table 4.

Revision of the Directions. Since many children experienced difficulties in remembering the names of the characters in recall, a slight change was made in the instruction supplied when the child was asked to recall the story, so that the names of the characters were now given:

> Now, could you tell me the story you read about...
> (insert names of story characters).

The instructions given prior to story reading were left unchanged.

TABLE 4

Example of two revised stories of the ARP.

FORM A — The Ball — LEVEL 1

Dan and Kim sat on the grass. It was a nice sunny day. But they were sad because they had lost their new red ball. They wanted to find it, so Kim looked among the flowers and Dan looked in the trees. Dan looked up high and saw something in a tree. It was the ball. Kim helped him climb up in the tree. When he got the ball he threw it down. Then they were very happy. They had a good ball game and lots of fun.

FORM B — The Dog — LEVEL 3

One Saturday morning Susie and her brother Andy looked sadly out the window at the rain. They could not go outside to play and the T.V. was broken, so they had nothing to do. Their dog, Rex, was sleeping peacefully under the table. Andy got his jeans, his red shirt and a big hat. He went silently towards Rex and got one paw almost into the shirt sleeve. Then Rex jumped up barking. Susie had to hold him while Andy got another paw into the second sleeve. Rex was annoyed and ran off with the shirt. He hid under a bed, still wearing it. The red shirt got filthy dirty. They had to wash it before mother came home. They decided that dressing dogs was not such a good game after all.

Revision of Miscue Analysis. The revisions to the miscue analysis were made so as to (1) eliminate from the analysis categories of information, e.g., coding errors as high, average, or low attention to graphics, which lacked precision in the measurement of processes; (2) introduce new categories which would better differentiate between processes, and also add more depth to the analysis; and (3) establish an outside criterion for a child's overall comprehension of the author's intended message.

The revised miscue sheet is presented in Figure 5. The revised miscue categories will be explained in a manner similar to the explanation of the original sheet, here the various columns are numbered from 1 to 17, and the processes which may be inferred from each are marked directly on the scoring sheet in Figure 5. In Figure 6, may be found examples of how eleven miscues from Track and Field (level-five) are scored; these miscues are numbered and will be referred to during the explanation of the coding system. In addition, these miscues are marked on the passage in Table 5. As the scoring is explained, three categories of data will be supplied: the type of miscue information that is recorded in that column, directions for scoring, and which process or processes may be inferred.

The raw processing scores may be used as descriptive data to characterize the reader's processing of connected discourse. However, a proportion-of-processing score may also be calculated for the reader on each of the processing indicators.

Text	Miscue	Corrected + / -	Text word length in graphemes	Graphic cues accessed	Graphic cue position I M F	Cue sequence + / -	Text word length in phonemes	s/s association phonemes	s/s association position I M F	real/non word accessed RN	unit accessed Ph. W. Pa.	Oral language (reader's meaning)							Author's meaning		
												Syntax		Semantics			Prediction		No change	Minor	Major
												(a) Sent.	(b) Part.	(a) Sent.	(b) Part.	(c) Pass.	(a) Sent.	(b) Part.			
1	2	3	4	5	6	7	8	9	10	11	12	13		14			15		16	17	18
		Monitoring		Attending	Attending/analysing	Sequencing graphemes		Sound-symbol associating	Phonemic chunking/analysing/sound associating	Meaning associating	Analysing/meaning and sound associating	Syntactic predicting	Syntactic predicting	Sentence meaning synthesizing	Partial sentence meaning synthesizing	Passage level synthesizing	Sentence syntactic and semantic predicting	Partial sentence syntactic and semantic predicting			

Figure 5. Resived Miscue Analysis Sheet

Text	Miscue	Corrected + / -	Text word length in graphemes	Graphic cues accessed	Graphic cue position I M F	Cue sequence + / -	Text word length in phonemes	s/s association phonemes	s/s association position I M F	real/non word accessed RN	unit accessed Ph. W. Pa.	Oral language (reader's meaning): Syntax (a) Sent.	Syntax (b) Part.	Semantics (a) Sent.	Semantics (b) Part.	Semantics (c) Pass.	Prediction (a) Sent.	Prediction (b) Part.	Author's meaning: No change	Minor	Major
1	2	3	4	5	6	7	8	9	10	11	12	13		14			15		16	17	18
(1) the	--	--										+	+	+	+	+	+	+	+		
(2) sprinting	spurting	--	9	7	IMF	--	8	6	F	R	W	+	+	+	+	+	+	+			
(3) competitions	compet-i-tions	--								R	Pa.	+	+	+	+	+	+	+	+		
(4) --	their	--								R	Ph.	+	+	+	+	+	+	+	+		
(5) win even	even win	--	7	7	IMF	--	7	7	IMF	R	Ph,	+	+	+	+	+	+	+	+		
(6) jump just for fun	jumping for the fun of it	--	4	4	IMF	+	4	4	IMF	R	Ph.	+	+	+	+	+	+	+	+		
(7) routine	routaine	--	7	7	IMF	+	5	4	IF	N	W	+	+	-	-	+	-	-		+	
(8) sprinted	spr--in	--	8	5	IM	+	8	5	I	N	Pa.	-	-	-	-	+	-	-		+	
(9) ankle	angle	--	5	4	IMF	+	4	3	IF	R	W	+	+	+	+	-	+	+			+
(10) badly	hardly	--	5	4	MF	+	5	4	F	R	W	+	+	+	+	-	+	+		+	
(11) being	beginning	+																			
TOTAL	11	10																			

Figure 6. Examples of Errors Coded

TABLE 5

Example of coding oral reading errors. The key words are underlined.

FORM A

Track and Field

Frank, Mary and their friends are interested in track and field.
Frank excells at running and Mary at the high jump.. The friends
spurting
are not real champions but practice hard to improve their sprinting.
Last Spring they decided they would enter all suitable events in
com-pet-i-tions
the track and field competitions. Although they followed a strict
Routine
training routine for a month leading up to the competitions, their
plans did not turn out as expected. Mary and the friends practiced
their
their jumping every day and did exercises to strengthen their leg
muscles. Frank was determined to become the junior champion of
the one-hundred metre dash so it was important for him to move
instantly on the sound of the starting gun. Every day he practiced
his start by imagining himself pushed forward by a strong wind.
At last the big day arrived. Mary was the first to compete. She
expended all her energy but still only managed to get third prize
for jumping. Their friends did not win even a single prize. Now
all their hopes were on Frank to become the champion at the one-
hundred metre dash. But after the jumping events, Frank thought
ing the of it
he would try the high jump just for fun. When the area around
the jump was deserted, Frank sprinted, sprang high in the air
and cleared the bar. Unfortunately for him, he landed in the
angle hardly
sand crying with pain. His ankle was badly sprained. Mary and
the friends came running. They helped him to the first aid room
beginning C
just as the one-hundred metre event was being announced.

Since this proportional score is required for the factorial validation study, this scoring procedure will also be described below.

1. The first column contains a record of the text word or words (in the case of compound errors, e.g., miscue 6) which have been mispronounced or omitted. If the error is an insertion of a new word this category will remain blank as in miscue 4.
2. Column two contains an exact transcription of the text or insertion as uttered by the child. If the error is an omission, this column is left blank (see miscue 1).
3. This column records whether the miscue has been self-corrected or not; correction is an indicator of monitoring, as is evident from error number 11. Self-corrected errors are not coded on grapho-phonics, or syntax and semantics, as it was felt that these miscues did not represent the same processing as that of uncorrected miscues. However, it is recognized that insights into the processing of text can be gained from an analysis of the types of errors different readers correct (Beebe, 1980).

   A total proportional score for monitoring is obtained by dividing the number of self-corrected

miscues by the total number of errors, corrected and uncorrected.

4. Column four contains the text word length in graphemes or, in other words, the number of letters in the word. This is not processing data but it will be used to calculate attentional behavior in column five.
5. This column contains a record of the graphemes in a word the reader attended to as determined from his error. The error is compared to the text word, and the graphemes which are uttered in any possible pronunciation are accepted, e.g., when "slough" is pronounced "slow" the child is given credit for attending to all the graphemes, as /o/ is a possible pronunciation for the orthographic pattern "ough". The total score on attending to graphemes is calculated as a proportion of the total number of cues in the text words. When all errors are analysed, the total number of cues accessed in all errors is calculated and divided by total number of cues in corresponding text words (column four).
6. The sixth column contains a record of the word parts --initial, medial, and final--attended by the child in order to select cues. This information gives an indication of the type of analysing behavior the reader engages in.

This stage of the analysis is accomplished by dividing the text word into three parts -- beginning, middle, and end -- which are defined as follows:

ONE SYLLABLE WORDS

Beginning:	initial grapheme or grapheme cluster
End:	final grapheme or grapheme cluster
Middle:	remainder

TWO SYLLABLE WORDS

Beginning:	same as above except if initial syllable
End:	is a prefix, and final syllable is a suf-
Middle:	fix, that syllable is the beginning or ending

THREE SYLLABLE WORDS

Beginning:	first syllable
End:	last syllable
Middle:	remaining.

The miscue is then compared to the text word in terms of these parts; and if a grapheme from any one of these parts appears in the miscue, the child gets credit for analysing the word and attending to that part. Sequencing of letters in the miscue is not a consideration at this stage, e.g., miscue 2 on Figure 6.

A total score is computed by counting the number of initial, medial, and final parts attended to; then each sum is divided by the total number of miscues which are coded on graphics.

7. Information in this column indicates how the reader sequences graphemes. Here the coding procedure is

straightforward: a plus sign is entered for miscues without sequence change, and a minus sign when there is a change. Not all errors are coded on sequence, for instance: omissions (miscue 1), insertions (miscue 4), oral errors coded only for partial accessing (miscue 3), and miscues bearing no graphic similarity with the stimulus word, e.g., "a" for "his" or visa versa.

Total scores are computed by dividing the total number of miscues with sequence changes, and the total number of miscues without sequence changes, by the total number of miscues coded for sequence.

8. The eighth column contains a record of the number of phonemes in the text word for all miscues which are coded on graphics; this is not in itself an indicator of process, but will be used when calculating the number of phonemes attended to in column nine.

9. The digits in this column indicate the number of individual phonemes of the text word pronounced in the miscue. This is computed by comparing the miscue to the stimulus word in column one in terms of phoneme similarity. A total score is calculated by dividing the total number of phonemes correctly accessed by the total number of phonemes in the stimulus words (column eight). This score is an indicator of the amount of <u>sound-symbol association</u>

the reader is engaging in.

10. This column was designed to yield information concerning three distinct processes: analysis, sound-symbol association by position and phoneme synthesis. Similarly to column six, this column also yields three categories of information. The miscue is compared to the stimulus word, and a decision is made as to how many of the parts -- initial, medial, or final -- are pronounced correctly and completely. Many of these units are syllables, hence information is gathered concerning phoneme synthesizing.

    A final score is calculated by dividing each of the three totals - initial, medial, and final - by the total number of miscues coded on graphics.

11. The eleventh column lists the number of word cues that have been associated with meaningful response, thus giving an indication of meaning association. Both the numbers of meaningful lexical items and pseudo-words pronounced are recorded. Real words are coded as "R" and pseudo-words or fragments as "N" (non-word). The two final scores are calculated by dividing the sums of "R's" and "N's" by the total number of miscues coded on graphophonics plus insertions.

12. The data in column twelve gives some indication

about the type of phonological or meaning unit the reader accessed in response to the visual cues he selects. Hence, information may be inferred about <u>attending</u> and <u>analysing</u>, in addition to information on <u>sound-symbol</u> <u>association</u> and <u>meaning-symbol</u> <u>association</u>. Three types of unit accessing are coded: "w" is entered for each word unit produced (pseudo or real) in response to the graphics; "Pa" is entered when an attempt is made to access a word through analysis, although this is only perceivable when the response is incompletely synthesized or blended. Word fragments are also indicative of partial accessing. And finally, "PH" is coded whenever there is an indication that a complete phrase unit is accessed. Typically this occurs in compound errors and is often accompanied by omissions and insertions. Examples of these three types of coding are given in Figure 6.

Total score are computed by dividing each of the three sums, "W", "Pa", and "PH", by the total number of errors coded in that column.

13. The degree to which the child uses <u>syntactic prediction</u>--at the sentence or part-sentence level--is coded in the two columns 13a and 13b.

If the miscue is syntactically acceptable at the sentence level in terms of oral language, a plus

sign is entered in 13a; if not, a minus sign is entered. At this phase of the analysis emphasis is on form, not on content: nonsense may thus be coded as acceptable, provided it sounds like English. When the miscue is syntactically acceptable at the sentence level, it automatically follows that it is acceptable at the part-sentence level. But if a minus is entered in column 13a, the error may form an acceptable syntactic pattern with the preceding or following sentence context, and thus a plus is entered in column 13a. In order to make these judgements, the error is read with all the other uncorrected errors in that sentence or partial sentence.

Total scores on _sentence_ or _partial syntactic prediction_ are obtained by dividing the sum of plus scores by the total number of uncorrected errors.

14. Columns 14a, 14b, and 14c contain an indication of _semantic synthesis_ at three levels: sentence, partial sentence, and passage. First, the sentence is read aloud with all the uncorrected errors included. If it has some meaning regardless of passage meaning, a plus is coded in 14a and 14b (partial sentence). On the other hand if the error is not semantically acceptable at the sentence level, the error is read twice: once with preceding context

and once with the following context; then if it is meaningful in either of these contexts a plus is entered in 14b. As in the case of syntactic acceptability, sentence and partial sentence context is defined as that unit of text where all of the reader's uncorrected errors are included.

Passage level semantic acceptability is judged on the compatibility of the error with the gist meaning of the story as has been read so far, e.g., on the passage entitled Budgies. If every time the child reads "budgie" or "budgies'" he reads "billing bug" or "billing bugs" and proceeds to make other appropriate changes in the story, e.g., reading "caged bugs" instead of "caged birds", these errors are acceptable in terms of the reader's meaning because they are appropriate for the schema he employs to build a coherence network. Such a reader is deriving meaning from print, but since it is not the author's meaning, he is not comprehending the printed message.

It should be noted that an error may not be acceptable semantically at the sentence or part-sentence levels, but nevertheless is acceptable at passage levels because these errors do not contradict key ideas within the story as it is conveyed in oral reading. This is typically the case when the child

produces a pseudo-word for a minor concept, e.g., miscues 7 and 8 on Figure 6.

Total scores on the three semantic columns are calculated in exactly the same way as the total proportions of syntactic acceptibility in 13a and 13 b.

15. Data in colums 15a and 15b give an indication of how the reader is utilizing <u>both</u> syntax and semantics simultaneously as an aid to constructing the authors meaning, thereby yielding information on <u>predicting</u>. A plus sign is coded in column 15a and 15b if the error is syntactically <u>and</u> semantically acceptable at the sentence level. If an error does not indicate prediction at the sentence level, it may do so at a part-sentence level, provided that error is both syntactically and semantically acceptable at the part-sentence level.

Total scores on both levels of prediciton are calculated in exactly the same manner as the totals in 13a, 13b, 14a, 14b, and 14c.

16-18. Unlike the scores collected in the previous columns, which were designed to yield information on processing, the data in columns 16, 17. and 18 act as one of the outside criteria for both the reader's constructive and reconstructive processings. Columns 16, 17, and 18 respectively indicate whether

no change, minor change or major change occurs in the author's meaning as a result of miscue, and tell to what extent the reader is comprehending the author's intended meaning.

The following criteria are used to judge the degree of changes in the author's meaning.

No Change: Errors which do not change the author's intended meaning in any significant way, e.g., changes in function words such as articles: "he saw horses in the field" for "he saw the horses in the field"; changes in adverbs such as "when the sun was rising" for "as the sun was rising"; or changes in word sequence: "There is it" for "There it is". However, each judgement must be made within the context of the entire text meaning, up to and including that error; an insignificant miscue in one context can be highly significant in another.

Minor Change: Minor meaning changes occur when the error changes the meaning of the sentence in which it occurs but not the gist meaning of the entire passage. In the present study, however, a more decisive criterion was devised: the key content words, the most important one-third of the passage, (nouns, verbs and some adjectives) were computed; then, if the error changed the meaning of a sentence but did not change a key

word, it was scored as a minor meaning change.

Major Meaning Change: An error is scored as a major meaning change when it significantly changes the meaning of the key words in the passage. The key words are underlined for Track and Field on Table 5.

Total scores for the three types of meaning change are calculated by dividing the sum of scores in each column by the total number of uncorrected errors.

Revision of Recall Categories. Only two minor changes were made to the recall analysis; two of the categories -- text experiential and text erroneous -- were subdivided. These categories contained information pertaining to two distinct processing strategies. Text experiential information (D) was divided into text experiential: instantiating-inferencing (D1) and text experiential: elaborating (D2). These two categories divide the information that was formerly contained in the one experiential category. Now the instantiating-inferencing category (D1) contains a record of recalled information that is constrained by the text; and the elaborating category (D2) is used to record experiential intrusions or storyline additions, and is merely triggered off by the theme of the text. The additions in this category are not considered as inferences, since they are not constrained by any specific part of the text. This revision was prompted by a suggestion of Kavanaugh (1981).

Text erroneous category was split into text erroneous:

associating (E1) and text erroneous: synthesizing-inferencing (E2). Whereas formerly all examples of faulty processing were coded in one erroneous category, now the different types of faulty processes are separated. As the names suggest, text erroneous: synthesizing-inferencing (E1) includes faulty synthesizing of ideas and erroneous inferences.

A sample recall protocol produced following the reading of Track and Field is presented in Table 6. This recall is analysed according to the system of Fagan (in press), and includes the revisions detailed above.

A final score for each recall category may be calculated by either of two methods: (1) the clauses scored in each category may be computed as a proportion of the total number of clauses recalled by that child; or (2) clauses in the various categories may be presented as a proportion of the total number of clauses in the stimulus passage. At present, it is not known which scoring method is the more sensitive measure of processing. Hence it was decided to include both measures in the factor analytical study to see if they behaved in a similar manner, viz., were they highly correlated and did they load on similar factors. If this indeed were the case, either measure could be employed in a study based on correlational data ; however, it should be borne in mind that this may not be suitable for studies which compare mean performance or absolute quantity of recall.

TABLE 6

Sample protocol including coding into recall categories.

Track and Field

Frank and Mary are [A] interested in track and field// and they
to do the competitions and [C] be champions in the next track
and field/ <u>join [B] all events</u>// they were [B] practicing every day//
their friends were [B] practicing all the time doing special
exercises // Mary was [D] practicing to do her high jump// and
[B] then the big day came// and she [B] won third prize// by the end
of the day [E2] they all won prizes// then Frank wanted to go high
jumping over that bar [C] just for the fun of it// after he [E1] wasn't
hurt// but they took him to [B] the first aid room// (then the nurse
no first and then) and then [B] the hundred metre dash was announced//
so he couldn't run [D1] in it// so the nurse [D2] gave him a check-up
instead.

Conclusion

A revised model of test construction, based on the work of Wardrop et al. (1978), was employed in the construction of the ARP. The major departure from conventional practice in test construction was to conceive of item generation and revision as a part of construct validity. This was both technically expedient and conceptually plausible because: (1) Construction of the scoring system involved the use of the theory of the assessment of reading processes to infer hypothetical constructs (processes); (2) An analysis of the inferential steps involved in moving from product to process is a major step in judging whether the ARP does indeed measure reading processes; and (3) Other forms of validity, e.g., content validity, may be considered part of the evidence for establishing construct validity. Therefore, an element of construct validity may be claimed based on the criteria which were specified for passage structure and content.

In the following chapter the factor analyses data will be presented, and the validity of the ARP will be further discussed in terms of this information.

CHAPTER VI

PROCEDURES: FACTORIAL VALIDITY

Introduction

The purpose of this chapter is to describe the study which was carried out primarily to collect factorial evidence to explain some links in the nomological network underlying the construct of reading as cognitive and linguistic processing. As noted in the previous chapter, this study is part of the ongoing process of establishing construct validity for the theory of assessment of reading processes and the ARP, the instrument devised to measure these processes.

Design

A study was designed which would highlight links or interdependencies between different dimensions of the interaction between reader and text. In the theory of the assessment of reading processes, this interaction was explained as cognitive and linguistic processing. Since at present, the measurement instrument designed to assess eight of these processes gathers data on fifty-four processing indicators, this study should employ a technique of investigation to discover the underlying dimensions which determine various types of text processing. A factor analysis provides such a tool. As Cattell (1952) states: "Factor analysis, carried out on the correlation coefficients, shows us how some variables can be grouped together because they behave in the same way, and it proceeds to delineate

new independent, underlying factors which may be responsible for these groupings" (pp.14-15). Moreover, factor analysis is widely accepted as a valuable technique in establishing construct validity; "one of the major aspects of the explication of constructs is in determining to what extent hypothesized measures of a construct measure the same thing, or break up into clusters of variables that measure different things" (Nunnally, 1978, p. 329).

Once it was determined that a factor analytic technique would be used to investigate inherent clusterings within the data, certain restrictions were imposed on the design, especially with regard to sample size and type of variables included in the study. A rationale for the various decisions will be provided below.

The method of data gathering was structured according to components of the theory of the assessment of reading processes. Thus, a sample of 102 average grade four readers read and recalled four passages from the ARP--two were read orally and two silently. The type of processes the readers engaged in during reading and recall were inferred using the ARP's miscue and recall analyses. Information was also gathered on six non-processing dependent variables. Then a factor analysis was carried out to group the independent processing and dependent variables.

Questions to be Considered

> Factor analysis provides...a method far more free than most methods from the necessity to elaborate rigid hypotheses. It is the ideal method of open

> _exploration_ in regions unstructured by present knowledge. In embarking on a factor analysis one need not have any more definite idea than Columbus had of America in regard to what is to be found. It is sufficient to hypothesize that _some_ structure lies there (Cattell, 1952, p. 14, emphasis in the text).

Even though the present exploration will proceed without any definite hypotheses, perception of what will be constructed as meaningful clusterings is to some extent predetermined by theoretical speculations about the orderly nature of the theoretical domain. Certain broad areas of processing structures could be expected, but without any definite speculations as to the nature of these groupings. For example:

1. The indicators of word-processing may group; however, it was not known whether such a clustering would simply include grapho-phonic processing or whether it would also encompass aspects of meaning association. Furthermore, since some of the indicators of graphic analysis and meaning association, such as partial accessing of words and producing meaningful words, were used for the first time, the speculations as to where they would actually group was open.
2. The indicators of syntactic and semantic processing should group; but the specifics of the clustering were completely unformulated. Indicators of such processing at various levels--word, partial sentence, sentence and passage levels--could have grouped

on their own or in various combinations.

3. Various aspects of discourse processing over the two modes of reading could have been expected to form groupings: indicators of text association, indicators of text synthesis, indicators of reconstructive processing and indicators of erroneous processing.

At that stage of theory development only these very general speculations were warranted. No speculations were put forth as to how the dependent variables -- standardized reading comprehension, degree of meaning change, and number of miscues -- would relate to processing variables. Nor was it possible to speculate on the nature or the degree of correlation between word-processing, syntactic and semantic processing, or aspects of these.

Variables

In a factor analysis there is generally no distinction between dependent and independent variables. Because in the present validational study there was a need to set up some outside criteria for processing behavior, some dependent variables were introduced into the assessment battery.

Dependent Variables. Three of the dependent variables have been described earlier when delineating the various phases in the construction of the ARP. These are: no change in the author's meaning, minor change, and major change. The fourth dependent variable is also based on oral reading errors, and

consists of the number of uncorrected errors a child makes when reading the passages as a proportion of the total number of words in those passages; whereas the fifth dependent measure concerns the level of the passages which were accepted for analysis.

The sixth and final dependent measure included is a grade-equivalent score from the comprehension subtest of the Canadian edition of the Gates-MacGinitie Reading Tests (G-M-R). This standardized test is administered on a system-wide basis in June each year to all grade three level students in the Edmonton Catholic Schools. Hence, this was obtained at the end of the students' year in grade three. The score indicates the students' ability to read short descriptive passages silently, and then to select among four multiple-choice responses in order to answer questions and to complete sentences.

The 1978-79 Canadian norms for the twelve levels of the G-M-R were developed from the results of testing 46,000 students in November 1978. Reliablity coefficients were computed from the standardization population for each level of the test -- between 3,000 to 4,500 pupils at each grade level. A Kuder-Richardson Formula 20 coefficient of 90 is given for Form C1 (grade three). In the teacher's manual of the test, appeals are made to content validity by referring to the sampling plans used for choosing the content and levels of the comprehension questions (literal or inferential) included. Furthermore, teachers are admonished "to look carefully at the test with (their) curriculum in mind" (MacGinitie et al., 1978, p. 23).

Initially, it was planned to include some other independent variables; however, owing to certain limitations put on the type of variable which may be included in a factor analysis, these were eliminated. Variables bearing a direct linear relationship with any of the other variables may not be included. Because of this, four variables were eliminated from the variable pool: proportional length of recall given after oral and after silent reading, and the proportions of these two recalls which indicated comprehension of the author's meaning (sum of text exact (A), text specific (B), text entailed (C) and text experiential: inferred (D1) in each recall). Both these sets of variables are composites and are made up of the sum of the other discourse processing indicators.

Independent Variables. Fifty-four independent variables were included in the pool of observations which were collected. These included fifty-three indicators of processing and a measure of intelligence.

The processing variables are taken directly from the miscue analyses which were introduced in the sections on item generation and revision in the previous chapter. Scores on twenty-five independent variables were taken from the miscue analysis. These were:

1. proportion of miscues corrected
2. proportion of omitted word in oral reading
3. proportion of inserted words in oral reading
4. proportion of graphemes attended to

5. proportion of initial part of word attended to
6. proportion of medial part of word attended to
7. proportion of final part of word attended to
8. proportion of graphic cues correctly sequenced
9. proportion of graphic cues mis-sequenced
10. proportion of grapheme-phoneme association
11. proportion of initial or word with correct grapheme-phoneme association
12. proportion of medial units of word with correct grapheme-phoneme association
13. proportion of final units of word with correct grapheme-phoneme association.
14. proportion of real words accessed
15. proportion of pseudo words accessed
16. proportion of word units accessed
17. proportion of partial word units accessed
18. proportion of phrase units accessed
19. proportion of miscues syntactically acceptable at the sentence level
20. proportion of miscues syntactically acceptable at the partial sentence level
21. proportion of miscues semantically acceptable at the partial sentence level
22. proportion of miscues semantically acceptable at the partial sentence level
23. proportion of miscues semantically acceptable at the passage

level

24. proportion of miscues indicating prediction at the sentence level

25. proportion of miscues indicating prediction at the partial sentence level.

Scores on twenty-eight indicators of processing were taken from the recall analysis (Fagan, in press). Each recall analysis consists of information on seven categories of information included in a recall. These categories are:

1. text exact (A)
2. text specific (B)
3. text entailed (C)
4. text experiential: inferencing-instantiating (D1)
5. text experiential: elaborating (D2)
6. text erroneous: associating (E1)
7. text erroneous: synthesizing-inferencing-elaboration (E2).

Total scores on each category may be calculated in two different ways (see Chapter 5, p. 155). Since there is no guidance from either the theory of assessment of reading processes, or research as to which measure would be the best indicator of processing, it was decided to include both measures in the variable pool. This decision resulted in the total of twenty-eight indicators of discourse processing: fourteen based on the oral reading recall and another fourteen based on the silent reading recall analysis.

The final, independent variable was the full-scale

score from the Lorge-Thorndike Intelligence Test (L-T-I) -- a conventional measure of "intelligence". In the first chapter, Feuerstein's (1979) dynamic view of intelligence as a modifiable set of cognitive processes was discussed. Even though some of the processes which he identifies bear different labels, they refer to mental operations similar to those hypothesized to take place during reading. In contrast to many conventional measures of intelligence, Feuerstein developed the dynamic assessment (LPAD) to investigate the processes rather than the products of intelligent behavior. Although tests of the traditional variety measure the products of "intelligence" as a set of relatively independent abilities, yet in order to achieve a certain score an individual has to engage in cognitive processing (see Cattell, 1971, pp. 23-46, for a discussion of these abilities). The question then becomes one of determining the relationship between the reading processes and the cognitive abilities necessary to complete an intelligence test.

In previous research, scores on standardized reading comprehension tests and on intelligence tests were correlated. As Carroll (1972) testifies:

> Comprehension ability tests tend to be substantially correlated with 'intelligence' tests, even those of a nonverbal character, such as a figure analogies test...One possible source of this correlation is the fact that reading and listening comprehension tests do not measure <u>only</u> what may be called 'pure' comprehension of language; because of the way in which they were constructed, and the kind of items they include, they tend also to measure ability to make inferences and deductions from text content (p. 3, emphasis in the text).

In the light of this fairly well attested relationship, it was decided not to include an intelligence test score as a dependent variable along with the standardized comprehension score. Furthermore, at this stage it is not certain whether reading is dependent on "intelligence" or visa versa. Hence, the intelligence test score was regarded as an independent variable along with all the other processing indicators. Since "one of the many difficulties surrounding IQ tests is that no one really knows what intelligence is" (Montague, 1975, p. 3), there is considerable controversy about the exact nature of the cognitive operations or abilites that are tapped in these scores. But these issues are not of concern in the present work; it is enough to assume that some mental processes take place as the test is completed. What is of interest, though, is: to which of the reading processes, measured by the ARP, is such a score related?

The particular IQ measure used in this study, the Lorge-Thorndike, was selected because it is the only intelligence test administered to students in the middle grades in the Edmonton Catholic School system. This score is derived from the verbal and non-verbal sub-tests, which seem quite narrow in focus even though the test manual does claim that the L-T-I measures reasoning:

> The items for the Lorge-Thorndike Intelligence Tests were selected so that for the most part they deal with symbolic relationhsips. In answering most of the items a pupil is required to discover a principle and then apply it. The tests, then, have been

designed to measure reading ability" (Lorge, Thorndike, and Hagen, 1967, p. 29).

The L-T-I was administered by the classroom teachers to the entire class at the end of the first semester in grade four.

Sample Selection

In order to employ an exploratory factor analysis as a means of clustering variables, a case base of approximately one hundred is the minimum required. Furthermore, if groupings among the processing indicators did emerge they would have to be generally interpretable within the theory of reading processes. Although this theory postulates the occurrence of eight processes which take place as the reader constructs and reconstructs the author's meaning, little is known about how this processing changes across the various stages of learning to read. So, if a sample from several different grade levels were chosen, the result might not be interpretable due to different processing emphases which may emerge at stages of reading competence. Hence, a homogeneous sample was sought.

A sample of 109 average readers from grade four read and recalled the stories. These children attended six different schools in the Edmonton Catholic School system. Initially it was decided to choose students who scored between the fortieth and sixtieth percentile on the Gates-MacGinitie Reading Test. However, because of the difficulty of securing access to more schools, these limits were expanded to include the thirtieth and eighty-ninth percentiles. All children within these percentile ranks

were selected, but seven of these children were later dropped from the study because they either made too few or too many miscues. Hence, the final sample numbered 102 students.

Instruments

In order to use a miscue analysis children should make a certain number of oral reading errors; therefore, they should be asked to read passages at or above their grade level. For this reason, grade four students were asked to read passages from levels four, five and six of the ARP. Form A (sports topics) and Form B (animal topics) were regarded as a unit, as were Forms C (sports topics) and D (animal topics); children read passages from one form orally and the alternate form silently. Twenty-six children read Form A orally and Form B silently, another twenty-six read Forms C and D in a similar set of random alternations.

The twelve stories including titles were typed on plain white paper for presentation to the students. Two sample passages are included in Appendix A.

Data Collection

The data were collected, over a two-week period towards the end of the first semester of the school year, by the researcher with the assistance of three graduate students in reading. All three students were experienced teachers and had some experience with informal reading inventories. The researcher met with them as a group before the data collection, and explained the study's

goals and procedures; in addition, the researcher was present in the schools on the day each assistant started work on this project.

On the first morning of data collection in each school, the researcher visited each grade-four classroom and explained the task to the entire class, also explaining that only some names were chosen from the class list. Then at the time of the assessment the children came individually to a quiet room where tape-recording equipment was set up.

Upon entering the assessment situation, the child was familiarized with the task and encouraged to ask any questions. The necessity for tape-recording was also explained. Then the student was engaged in a practice oral reading and recall task.

Following the practice, each student was asked to read and recall two passages from three levels of the ARP. At each level, oral reading and recall always preceded silent reading and recall. Prior to reading a particular story, efforts were made through questioning to ensure that the children would utilize the appropriate interpretative schema when reading, e.g., Do you know what beavers are?

When asking a child to read a passage, the examiner handed the child the typed page while giving the directions from the ARP:

> I want you to read this passage about...(insert topic). Later I will ask you to tell me the story. Read it out loud/silently. You may begin now.

Immediately following the reading, the passage was handed to the researcher. The student was then engaged in a conversation on a topic unrelated to that of the passage, for approximately three minutes, after which he recalled the story orally when directed to do so:

> Now, I want you to tell me the story you read about ...(insert names of characters).

All the oral readings and recalls were tape recorded and later transcribed.

Data Analysis

Prior to coding either oral reading errors or recall protocols, a decision had to be made regarding which passages to include in the analysis. At this stage it was intuitively obvious that about a quarter of the students were reading the level-six passages at frustration level. However, the level six passages could not be dropped from the study as twenty students would then have a very small number of uncorrected miscues. Hence, the decision was reached to include a sample of four passages, two oral and two silent, from each child's reading. The selection criterion was based on the number of oral reading errors, and therefore focused on the level of interaction between reader and text rather than on the level of text per se. On the two passage levels chosen for inclusion in the analysis, the reader should make a total of between ten and thirty uncorrected oral reading errors. The final choice was made in favor of the two

passages whose combined errors were closest to fifteen (the mean). As a consequence of this selection strategy, levels four and five were selected for the largest group of fifty-four, five and six for a group of forty-three, and four and six for the smallest group of six. The average number of oral reading errors for the 102 selected readers was seventeen. Seven students were dropped from the study at this stage: two because they had fewer than ten errors, and five because they made over thirty uncorrected miscues on any two passages.

All the oral reading errors were written on miscue sheets and coded according to the scoring rules of the ARP. Total proportions of the various types of processing indicators and dependent variables were computed. At the same time, each child's oral and silent reading recalls were analysed according to the system of Fagan (in press), and then coded in the seven recall categories. The total scores for both types of recalls were calculated in the two ways described earlier: as a proportion of total recall, or as a proportion of the number of clauses in the input passages.

Finally, the data from the miscue sheet, the recall analyses, the level of passages read, and the intelligence score were entered in the computer for statistical analysis.

Reliablity

The researcher met every week over a three-month period with her dissertation supervisor to discuss the changes that were

being made to the scoring systems. When an innovation was agreed upon, each would code a child's protocol independently to insure that the criteria for scoring were clear.

When an agreement was reached on the final version of the miscue and recall analyses, the researcher and her supervisor independently scored the miscues and recalls of eleven children. Inter-rater agreement was then calculated by employing the Arrington Formula as outlined by Feifel and Lorge (1950): the number of agreements between observers is doubled and divided by this product plus the disagreements, i.e.,

$$\frac{2 \times \text{Agreements}}{(2 \times \text{Agreements}) + \text{Disagreements}}$$

This formula produced the following agreement scores: .96 on the miscue analyses, .97 on the three categories of meaning change, and .95 on the recall categories. These scores were judged adequate to insure consistency of measurement.

Statistical Analysis

In order to cluster the variables and to explicate underlying structure, a correlation matrix was computed and was then factor-analysed using two rotational techniques: Varimax and Promax. Eventually, an eight-factor oblique solution was accepted for interpretation. The rationale for the steps involved in the analyses, plus an interpretation of the factors, will be given in the following chapter.

Conclusion

The procedures involved in the data collection and initial analyses of a factor analytic investigation were presented above. This study of the inherent structure of the construct of reading froms a part of the continuous process of construct validation of that hypothetical construct.

CHAPTER VII

FINDINGS AND INTERPRETATION

Introduction

"A factor problem starts with the hope or conviction that a certain domain is not as chaotic as it looks" (Thurstone, 1947, p. 55), and then proceeds to explore the latent structure in the data. This chapter will present an account of the extraction and interpretation of the factors. Since factors focus on the underlying construct rather than on specific loadings, where appropriate, examples typical of the underlying construct will be given. The implications of each factor for reading and identifying specific processes will be presented in Chapter 8.

Factor Analyses

The initial step in uncovering latent structure within the data was to calculate a correlation matrix (60x60) consisting of all the variables which were to be factor analysed. In investigating this matrix, two things were noted: (1) each variable was correlated significantly (.4) with some other variable or variables; and (2) the two different measures of each recall variable were always correlated above +.7, e.g., text exact oral calculated as a proportion of total recall correlated .890 with the same category of text computed as a proportion of the total number of clauses read. On the basis of these observations, it was decided to include all sixty variables in the factor analysis. The procedures used to determine the factors are considered to be

a mixture of what Horn (1967) terms subjective and analytic procedures. In subjective procedures, rotation is effected in such a way as to make the solution "meaningful"; whereas in analytic procedures rotation is based upon explicit mathematical conditions (p. 813).

The correlation matrix was analysed by a Varimax principal components technique. Since this was an exploratory factor analysis, the number of factors was not specified. Initially, eighteen factors, with eighteen eigen values greater than one were extracted and rotated to simple structure. These factors accounted for eighty percent of the total variance, but this solution was considered too unwieldly for the purposes of the present study namely to ascertain if the observed differences in reading performance could be accounted for by a smaller number of theoretically valid processes or combination of processes. A solution was, therefore, sought where a large amount of the variance among processing indicators would be accounted for by a smaller number.of factors.

To this end, the correlation matrix was then re-analysed by the principal components technique, successively extracting from seventeen to two factors. Each solution was orthogonally rotated by the Varimax method. The problem was then one of choosing among these seventeen solutions. There are few guidelines in the literature to assist the researcher in making this rather arbitrary choice. Kim and Meuller (1978) maintain that adherence to the "postulate of parsimony" offers a partial solution to this

dilemma. This principle urges the researcher to accept the "simplest" factor structure that accounts for the variance within the data. However, "simpler structure" for one researcher with a specific purpose in mind may not constitute a "simpler structure" for another with a different goal. In practice, as Nunnally (1978) demonstrates, the "number of factors problem" must be approached from two stances. From the statistical viewpoint, a good solution

> explains as much as possible of the variance h^2 of the variables. From the standpoint of empirical research, a good (solution) is one that is easily interpreted and/or relates most clearly to psychological theories (p. 346).

All of these considerations guided the selection procedure, but two additional criteria specific to the present analyses were also added: (1) During the initial data analysis and coding, the researcher observed some patterning of relationships between certain variables; hence, where possible factorial solutions highlighting these relationships should be chosen. (2) The two different scalings of each recall variable should load on the same factor and in the same direction. Balancing these criteria against each other, an eight-factor solution was chosen.

When the orthogonal eight-factor solution was satisfactorally interpreted, the factor loading matrix was rotated to an oblique solution by a Promax rotation to discover if these were in fact independent factors. This eight-factor oblique solution was selected for final interpretation, as it conformed to all the criteria related to elegance of structure and interpretability outlined above; and furthermore, having been subjected to a Promax

rotation, the relationships among factors could also be observed. The loadings of the sixty variables on the eight factors plus the communalities h^2 are presented in Appendix B. These factors accounted for fifty-six percent of the total variance.

For clarity of discussion only one set of the recall variables--the categories as a proporion of the child's total recall--will be referred to in the interpretation; this method of analysis has been employed by other researchers utilizing recall categories (e.g., Beebe, 1981; Brake, 1981; and Brailsford, 1981). The factors are identified on Table 6; in addition, the percentage of the total and common variance accounted for by each is given. The intercorrelations between the eight factors were low: only one value was above .3, and a further three were over .2. Factor eight: Error-Elaborating has negative correlations with three other factors: a) with Factor One, General Achievement — Monitoring (-.268); (b) with Factor Four, Semantic Constructive Processing Sentence Unit (-.245); and c) with Factor Seven, Semantic Constructive Processing — Transforming (-.35). Finally, Factor One, General Achievment — Monitoring, correlated .209 with Factor Two, Grapho-Phonic Processing. Presentation of the factors will not be ordered by size of variance accounted for, but rather by pragmatic considerations; they are sequenced in terms of the contiguity of the units of text processed.

Table 6. The Eight Oblique Factors (Promax) and Their Percentage of Total and Common Variance

	% common variance	% total variance	Rank in terms of variance accounted for
Factor 1			
General Achievement-Monitoring	15.5	8.4	second
Factor 2			
Grapho-Phonic Processing	12.1	6.6	third
Factor 3			
Unit Association-Sequencing	11.0	6.0	sixth
Factor 4			
Semantic Constructive Processing-Sentence Unit	19.4	10.6	first
Factor 5			
Constructing-Reproducing	8.2	4.5	eight
Factor 6			
Reconstructing-Synthesizing	9.7	5.3	seventh
Factor 7			
Constructing-Transforming	12.0	6.5	fourth
Factor 8			
Error Elaborating	12.1	6.6	fifth

Factor Solution

Factor One: General Achievement - Monitoring

Factor one appears to reflect an underlying behavioral dimension representing the grade-four students' level of scholastic achievment and monitoring during reading. The variables or processing indicators loading on this achievement-monitoring dimension are presented on Table 7.

Table 7

Factor One: Factor loadings greater than .4 from Oblique Rotation (PROMAX)

Standardized Reading Comprehension (GMR)	.650
Level of Passages Read	.650
Intelligence Test (LTIF)	.474
Errors Corrected	.465
Partial Words Accessed	.460
Uncorrected Errors	-.867
Text Erroneous: Synthesizing-Inferencing Silent (E2)	-.652
Text Erroneous: Associating Silent (E1)	-.463

A comprehension score on a standardized reading test is the highest loading on this factor. Unfortunately, there is an ongoing debate as to the meaning of widely used educational measurement concepts such as achievement, competence, mastery, and scholastic ability (Messick, 1981, p. 18). In general, educational achievement tests measure an examinee's mastery over a specific area of skills or curriculum content. The skills and content are supposedly hierarchically structured, often based on some form of task analysis. An operational definition of achievement--what

these tests measure--is definitely inadequate to convey the full sense of academic achievement. Nevertheless, scores on these tests indicate, at the time of testing, that children have developed certain skills in handling verbal, quantitative and symbolic information which contribute to their success in school work.

The achievement measure, the Canadian edition of the Gates-MacGinitie Reading Tests (GMR), fits in with the general category described above. In Level C of this test, the child is presented with short narrative and descriptive passages to be read silently followed immediately by multiple-choice questions focusing on details of the passage, drawing inferences, or recognising the main idea. Since the questions are all in a multiple-choice format, recognition rather than production is required of the child. Furthermore, since the answers to the twenty-two passages are scored dichotomously, exactness wins a high score for the child. In addition, these passages cover a range of difficulties; it then follows that the more able and exact readers get the highest scores.

Hence, it is hardly suprising that the level of passages read on the ARP is part of the same factor as the score on the Gates-MacGinitie; grade-four children who were able to read more difficult passages and answer comprehension questions on the GMR could also read the experimental passages written for grade-six with relative ease (as measured by uncorrected errors).

The third highest positive loading of the achievement

monitoring dimension is the score on the <u>Lorge-Thorndike Intelligence Test (LTI)</u>. Rather than measuring "innate" ability, IQ tests simply measure learned skills or scholastic achievment. Nevertheless, the GMR and LTI cannot be seen as alternate forms of one and the same test. Whereas only the reading comprehension score from the GMR was included in this study, a full-scale score from the LTI was included, which was based on several aspects of the vast range of achievement including verbal: e.g., vocabulary, reading and computation, and non verbal skills such as figure analogies. Thus, difference in range may partially explain the discrepancy in the two achievement loadings (.707 and .474) on this factor. In addition, there was a six-month interval between the administration of the two tests.

Correctional behavior or monitoring is the only specified indicator of processing to load positively with achievement, thus representing the higher achievers' tendency to engage in more self-corrections when reading orally. This finding is in agreement with other evidence based on miscue data: (1) Proficient readers focus more on semantic considerations whereas the less proficient focus more on graphics when reading (see Leu, 1982, for a review of the research). (2) Better readers, in terms of achievement, engage in more correctional behavior when the meaning of the passage is disrupted (e.g., Beebe, 1980; and Clay, 1969).

Yet, Biemiller (1970) cogently shows in his longtitudinal study that the more able grade-one reader learns to attend more closely to the graphics while monitoring meaning. At the

beginning of the year in the first grade, chidren's errors were constrained by preceding context, but by the end of the year the better reader's errors were both graphically and contextually constrained. These results were corroborated by the work of Clay (1969) in New Zealand. Overviewing her longitudinal research with first and third graders, Clay (1979) reports that over the first years at school children's awareness that their oral response is incongruous with the print develops steadily right from the initial phases of learning to read. Moreover, the child learns to monitor himself from a variety of cues including visual perceptual, morphophonemic, syntactic, and semantic cues (p. 153). But the reasons why the high achievers in this study tended to engage in monitoring cannot be resolved unequivocally at present, as the quality of the uncorrected errors has not been analysed. Yet, regardless of the categories of errors corrected, monitoring doubtless facilitated their obtaining higher scores on the GMR and the LTI, in addition to reading higher level passages on the ARP.

The fact that partial word accessing loaded positively on the same factor as self-correction may give some indication of the nature of high achievers' monitoring. Partial word accessing probably represents the children's attempts to match their response in an exact manner with units of print by analysing the graphics within a word. In the present study, there are many examples of children who read the level-six passages with relative ease but who stalled on the more difficult words (longer

and less frequent). For example, a child meets "enthusiastically" and reads "en-thus-i-as-tic-al-ly", with a slight hestitation between syllables. These children are monitoring graphics or more exactly sound-symbol associations each step along the way.

So far, this factor appears to determine a behavior complex where the reader searches for meaning, strives for exactness and is aware of the correct form possibly from both a semantic and a perceptual standpoint. Because of the dichotomous scoring of the GMR and LTI, exactness is of major significance towards achievement. But monitoring is no doubt another major contributing factor, where the higher achievers probably look back at the item or passage to confirm the answers. In a study on spontaneous lookbacks to resolve comprehension obstacles, Garner and Reis (1981) observed that good comprehenders and older readers engaged in more monitoring and look-backs when they failed to answer comprehension questions. Furthermore, Williams and Clay (1973) report that successful third-grade readers monitored 40% of all correct responses while reading orally. Monitoring in that study was gauged by word repetitions and hesitations which interrupted the flow of reading.

Moreover the negative loadings give further testimony to the foregoing characterization of the achievement-monitoring factor by indicating the type of behavior or processing that is dysfunctional for this dimension of reading: making a higher proportion of uncorrected miscues and recalling erroneous text material following silent reading. The tendency of proficient

readers to make few word identification errors is in agreement with the findings of Perfetti and Hogaboam (1975) which demonstrate that good comprehenders are able to identify words more quickly than poor comprehenders.

Furthermore, the fact that the higher achievers produced fewer faulty reproductions, transformations, syntheses, and inferences is hardly suprising. This demonstrates a faithfulness to the text, in contrast to the type of reading which is more inclined to take liberties with the exact message. Recalling inexact units of text may or may not be caused directly by word identification problems, as examples from protocols illustrate: (1) One child miscued on the word "bridle", which he read a "bridge" in the sentence "He slipped on the bridle". In his recall, this child reproduced the incorrect information due to miscue. (2) On the other hand, another child attended correctly to the words in the sentence "Their uncle gave them one yellow and two blue budgies", but in his recall substituted the word "dad" for uncle, thus reproducing erroneous information. These examples, however, come from oral reading and recall, and the erroneous processing being discussed is an indication of faulty processing during or following silent reading. Hence, a relevant question is: what is the relationship between oral and silent reading? This is a pertinent issue, since the indicators of achievement and monitoring in the two modes of reading are truly intertwined on this dimension; for instance, the reading involved in the GMR is silent, and the level of passages read is based

predominantly on oral reading.

Levin (1979) devotes a chapter to discussing this problem and reviewing relevant research. He concludes that although certain differences exist (e.g., speed of reading) the "central processes" in both modes of reading are similar (p. 37). The works of Beebe (1980, 1981) and Eagan (1973) support this conclusion. Beebe found that semantic and syntactic acceptability of miscues, self-corrections, and retelling based on oral reading predicted success in silent reading. As Beebe (1981) herself states, "the covariations in understanding in the two modes of reading are equally affected by the same predictors" (p. 83). Similarly, Eagan (1973) discovered that at the second- and third-grade levels the number, length, and placement of pauses during oral reading was related to comprehension in silent reading. On the other hand, Brake (1981) demonstrated that the recalls given by second grade children indicated a higher level of comprehension following silent reading, but that they recalled more erroneous material following silent reading (p. 67). All in all, this research asserts that, in general, oral reading may be used to predict silent reading as similar processing may be assumed to occur during both modes.

One final topic will be introduced in the interpretation of this factor, viz, the relationship between metacognition or monitoring, background knowledge, and achievement. The growing body of research evidence in this area of awareness of cognitive processing shows that consciousness of cognitive functioning is

developmental, but nevertheless the more proficient student is more aware of blocks in comprehension and hence monitors his reading (e.g., Baker, 1979; Baker and Anderson, 1982; Brown, 1975, 1980, 1981; and Flavell, 1981). Furthermore, Brown et al.(1981) point out that a reader may fail to do this for two reasons, "inefficient application of rules and strategies and impoverished background knowledge" (p. 18). If a child has inadequate background knowledge or vocabulary, he will have problems monitoring his reading comprehension.

Children have to have an adequate store of background knowledge in order to read varied material in the GMR, to comprehend the tasks in the LTI, and to read the higher level passages which included less familiar topics and vocabulary. In addition, it is easier to identify a word correctly if the word has been heard before (Perfetti, 1975)-- thus facilitating the accessing of the longer, less familiar words in parts. Fagan and Eagan (1982), as a result of their analysis of the word recognition strategies of children who made gains and children who made no gains in a remedial reading program, strongly recommend that children's vocabulary be increased if their word recognition is to increase. They found that although the children who made gains began to use larger chunks of graphics, they were sometimes forced to produce pseudo-words when a real word was not readily available to them.

In conclusion, it would appear that this dimension representing both a search for meaning, exactness, and awareness

of the correct form is closely related to the possession of relevant background knowledge. Clay (1979) states that monitoring behaviors are a means "by which children (can) teach themselves, irrespective of the (instructional) program they are in" (p. 150). Hence, the achievement evident on this factor may be a consequence of the children's ability to direct and monitor their own learning and interaction with print.

Factor Two: Grapho-Phonic Processing

Factor two, Grapho-Phonic processing, unequivocally represents the manner in which readers use their knowledge of graphic analysis and sound-symbol association to pronounce words. It is evident from Table 8 that all eight variables associated with this dimension are indicators of attention to graphics, analysis, and sound-symbol association.

Table 8

Factor Two: Factor loadings greater than .4 from Oblique Rotation (PROMAX)

Attention to Graphemes	.756
Attention to Initial Graphics	.706
Attention to Medial Graphics	.652
Attention to Final Graphics	.539
Association of Phonemes	.767
Association of Initial Phonemes	.734
Association of Medial Phonemes	.514
Association of Final Phonemes	.508

The two general indicators of grapho-phonic processing

of words -- attending to graphemes over the entire word and awarding those phonemes their correct phonological value -- proved to be the highest loadings on this underlying configuration of processing. Therefore, attending to the graphics and associating correct sounds with graphemes are closely related. In addition the high positive loading of attending to graphics in the beginning, middle and end parts of the word reflects the reader's attempts to analyse units from across the word in order to select appropriate cues. Moreover, the high loadings on associating the correct sound unit to the initial, medial, and final word parts indicates that the fourth graders tended to analyse print into chunks rather than individual letter units to accomplish sound-symbol association.

This type of grapheme analysis by readers is substantiated by the findings of Fagan and Eagan (1982). The readers who made gains in a remedial reading program showed a greater tendency to analyse graphics, and also changed from making responses on the basis of single letter cues to a strategy of utilizing larger size (especially four letter) units when making word identifications.

The isolation of grapho-phonic processing as a relatively homogenous factor distinct from concerns of meaning or comprehension is in harmony with the results of two recent studies. Prior to carrying out a factor analysis on miscue data of grade four readers, Beebe (1981) investigated the correlation matrix consisting of graphic, phonic, syntactic, and semantic variables, and

as a result hypothesized that "graphic and phonic cueing strategies were so closely related that they were not separate variables but together formed a grapho-phonic strategy" (p. 110), and in addition that syntactic and semantic cueing strategies probably formed a syntactic - semantic cueing strategy. Moreover, there was a low negative correlation between the grapho-phonic and the syntactic-semantic variables. The results of factor analysis confirmed these predictions: grapho-phonic processing and syntactic- semantic processing loaded on two independent factors.

Haupt and Goldsmith (1982) factor-analysed the miscue data from a comprehensive coding system (including graphic, phonic, lexical, syntactic, and semantic analyses). The results indicated that by the second and fourth grades, average reader's grapho-phonic processing represented a distinct factor from either syntax or semantics. Beebe (1981) reminds her readers that the emergence of grapho-phonic processing simply bears out the perceptions of professionals working with children learning to read:

> Many readers are overly concerned with accurate letter-sound relationships. Such an obsession with accuracy often leads the child to believe that as long as he reads as accurately as he can, that it matters little whether what he is reading...makes sense to him. Those readers who put...meaning first are usually those who are least concerned that their reading be an exact (or very close to exact) rendition of the text before them (p. 110).

Furthermore, Beebe (1981) found that the extent of grapho-phonic accuracy did not predict reading comprehension on a standardized test.

Smith and Holmes (1971) draw on research findings to

demonstrate the independence of letter, word, and meaning identifications by arguing that in order to reduce uncertainty in each case the reader draws on a separate set of distinctive features. Two recent studies by Fleisher et al.(1979) give further proof of the independency of word identification and comprehension. In the first study poor comprehenders were trained to decode isolated words; they then read a passage consisting of those words, but showed no gain in comprehension. Following this, the authors hypothesized that focusing on isolated words may have diverted the readers attention from considerations of meaning. So a second experiment was carried out, where poor comprehenders were trained to read words in phrase units. Yet phrase reading had no transfer to passage comprehension. Hence, Fleisher et al.(1979) conclude: "By themselves, these data suggest that if decoding speed is implicated in comprehension, the relationship may be one of necessity rather than sufficiency" (p. 47). And in a similar manner the pronunciation of words was a necessity in oral reading, rather than a sufficient cause of comprehension for the grade-four readers in the present study.

Finally, it should be pointed out that the readers did not necessarily devote conscious attention to decoding graphics. The process of attending, analysing, and sound-symbol association may have been executed automatically (Samuels, 1976), with no apparent loss in speed of reading. By the fourth grade the readers probably have accumulated a certain store of spelling patterns and a corresponding store of sound units in memory which can be

tapped without conscious thought.

Factor Three: Unit Accessing-Sequencing

The unit accessing-sequencing factor appears to represent a latent dimension isolating how grade four readers abstract units of graphics or meaning from the page, and how these units are sequenced in the oral production. Table 9 presents processing indicators loading on this factor.

Table 9

Factor Three: Factor loadings greater than .4 from Oblige Rotation (PROMAX)

Sequencing of Graphemes	.689
Word Accessing	.633
Faulty Sequencing	-.589
Phrase Accessing	-.583
Insertions	-.564
Partial Word Accessing	-.432

Whereas factor two reflected the type of processing the reader engaged in to pronounce word units, the present factor appears to reflect the combination of processing the reader engages in to access words, word strings, or pseudo words.

The two positive loadings for grapheme sequencing and whole word accessing indicates that the readers are inclined to engage in correct sequencing of the graphemes if they access words as intact phonological or meaning units. Probably what is involved in this complex of processing is a type of one-shot word-like association. An inference theory of word identification

best suits this phenomenon of contextual oral word identification. When a word is perceived, featural information is briefly held in memory. "Information in perceptual storage is then synthesized or categorized with respect to a set of relevant perceptual categories in memory" (Joula et al., 1979, p. 93). Because these perceptual categories may consist of word units (lexical items) or common orthographic patterns, a response may be inferred from the recognition of a few letters. Alternately, "The identification of a few letter sounds could (have been) used to generate the entire phonemic code, of the word, or at least a recognizable approximation to its normal sound pattern" (Joula et al., p. 94).

Evidence provided by Fagan and Eagan (1982) may help to explain when the word-like response was likely to be a real word or a pseudo word. Although the remedial children who made gains in the tutoring program still produced both real and pseudo-words, pseudo-words were produced more often for the longer, less familiar words. Therefore, the authors conclude that the children produced pseudo-words when a real word was not readily available. Furthermore, when these children "encounter a word for which a real word response is not available, they rely more heavily on the grapheme units within the words" (Fagan and Eagan, 1982, p. 9). Even though in the present task the words were read in context, a similar process is likely to happen (except that sentence level prediction may also have played a role). Word recognition has to be accomplished at such a speed that not all the graphics can be analysed, and so a response just "pops" into the child's

consciousness (Adams, 1978, p. 50) on the basis of orthographic, phonological, and semantic information in memory. An example of two adjacent errors from Cross-Country Skiing, Level 5, may help to illustrate some points made above:

> occessional
> The silence was broken only by the occasional howling
> distance
> of a distant wolf.

The less familiar word "occasional" triggered off a pseudo-word and the more familiar "distant" the related noun "distance".

Inversely related to associating or retrieving correctly sequenced wordlike units (actual lexical items or pseudo-words) is a complex of processes including: the tendency to engage in faulty sequencing of graphic units, inserting words in the text sequence in oral reading, and analysing and associating graphics at the phrase and partial word levels. All but one aspect of this processing of print could be potentially explained by the theory of reading processes outlined in Chapter 3, that is, the implied relationship between partial word accessing and faulty sequencing of graphic units. One would expect that a reader who engages in partial word processing is sequencing the graphemes singly or by unit. However, an investigation of the initial correlation matrix (60x60) helped explain this anomaly. Partial word accessing and faulty sequencing, in fact, are uncorrelated (.021), yet these load on the same dimension because of the strong relationships they bear to other aspects of word accessing and sequencing. An investigation, this time, of the protocols explained why accessing both partial words and phrases is inversely related

to word accessing sequencing. The protocols indicated that some children only show evidence of accessing graphics in word units while others accessed print using a combination of word and partial word or phrase units. This combination of processing probably indicates some type of processing flexibility. Furthermore, it was observed that the children who engaged in associating text in phrase units also engaged in mis-sequencing.

This observation is scarsely suprising since sequencing variables were coded for both word and grapheme sequencing, and the readers often mis-sequenced words when they accessed phrases because they were predicting syntax and meaning, e.g.,

Now they have only two...

"There it is".

In contrast, the missequencing of graphemes within word units was observed to occur for pseudo-words or when the children were predicting only syntax, e.g.,

obreserve their spaces forgin mitlary
TEXT to observe three species; foreign; military;

This type of pseudo-word faulty sequencing probably occured because the readers failed to associate the graphics to a familiar word or letter pattern.

This underlying processing dimension may be interpreted in the light of mediated and non-mediated processing (Smith and Kleinman, 1979). This issue concerns how the fourth-graders interpreted letter strings as words -- did they go straight from word features to words, or did they go through letters and letter sounds on the way to accessing words? The present dimension

represents immediate "word" identification because the graphics are immediately associated with a pseudo-word (sound) or word (meaning). Furthermore, this type of immediate word identification is accompanied by correct sequencing and their occurrence on this dimension may be further understood by reference to Smith and Kleinman's (1979) explanation of how readers can identify structured nonwords as easily as words: "The beneficial effects of structure on letter-string perception are due to the reader's gain in information about the order of letters" (p. 80).

Even though the loadings on this factor give some indication that the readers focused on word units to access meaning, phonological codes, or both, it is not known what precise features of the word were used to spark an association. But theorists are now beginning to speculate that readers may employ dynamic rules, not specific units, to analyse graphic input. Use of rules or heuristic strategies may explain the flexibility of processing evident in the processing complex that is inversely related to word accessing-sequencing; or as Smith and Kleinman (1979) assert: "It may be that (this factor represents a dimension where) people use both units and rules, although some rely more on units whereas others depend more on rules" (p. 81). Finally, Clay (1979) maintains that reading strategies are organized in different ways at different stages of learning to read (p. 157), therefore the inverse access and sequencing loadings may actually reflect processing at various stages along the route to reading proficiency.

In conclusion, this factor represents how the readers

gained immediate access to a word pronunciation with the same sequence as the graphemes on the page.

Factor Four: Semantic Constructive Processing-Sentence Unit

This factor reflects how the readers engage in meaning construction at the sentence, partial sentence, and word levels. The six processing indicators with significant positive loadings, plus the two with negative loadings on this dimension, are presented on Table 10.

Table 10

Factor Four: Factor loadings greater than .4 from Oblique Rotation (PROMAX)

Partial Sentence Predicting	.883
Sentence Predicting	.793
Partial Sentence Semantic Acceptability	.876
Sentence Semantic Acceptability	.815
Real Word Accessing	.852
No Meaning Change	.708
Pseudo-Word Accessing	-.885
Minor Meaning Change	-.503

When the readers were engaging in the strategies represented by these processing indicators, they were actively using their knowledge and experiences to interpret the graphic symbols. In a study with ninety-four average and above average grade-four children, Beebe (1981) found that "The better the reader was at utilizing his language knowledge in conjunction with his knowledge of the world in order to interpret pieces of incoming linguistic

information, the higher his comprehension score" (p. 185). Though in the present study the comprehension score on the standardized reading test did not load on this dimension, nevertheless, two other dependent variables representing possible changes in the author's meaning are present. Semantic constructive processing at the sentence level is related to the same underlying dimension as making oral reading errors that cause no change in the author's meaning, whereas miscues that result in a minor change in the author's meaning are negatively related to this dimension.

Authors who adopt a psycholinguistic theory of reading maintain that prediction is the key process in the construction of meaning (e.g., Goodman, 1976; Hockberg, 1970; and Smith, 1978). Furthermore, due to the influence of transformational grammar in the underpinnings of this theory, the focus of their explanations is on sentence-level phenomena i.e., they concentrate on syntactic and semantic prediction at the sentence level similar to the many variables loading on this processing dimension.

Sentence and partial sentence level prediction indicate how the readers predicted grammar and meaning as they read. In order to read at the optimum rate, the reader has to anticipate both the meaning and grammaticality of the sentences of print, and thus to reconstruct meaning using a minimum of graphic information. Since prediction was scored if an error was simultaneously syntactically and semantically acceptable, the prediction processing indicators may be said to subsume both syntactic and semantic acceptability at the sentence and part sentence levels.

The errors which were semantically constrained at the sentence and partial sentence levels are in evidence on this dimension, thus indicating semantic acceptability at these levels of text. But, contrary to expectations and to previous research findings, syntactic acceptability did not load on this dimension. Beebe (1981), for example, found that syntactic and semantic acceptability loaded on the same factor with the equivalent loadings of .863. Investigating the miscues of grade one children, Biemiller (1970) found that the majority of the errors were both syntactically and semantically constrained.

The obvious absence of any direct indication of sentence or partial sentence acceptability can possibly be explained by the fact that there are very few instances of syntactically unacceptable errors. The mean proportion of syntactically and semantically acceptable miscues are given below:

	Syntax	Semantics
Sentence	.900	.644
Partial	.980	.760

Very little of the variance in reading performance may be attributed to syntactic acceptability. This is substantiated by most previous research employing a miscue analysis. In an early study analysing 8,000 substitution errors (Clay, 1968) found a high instance of syntactic equivalence between the utterances of grade one children and the text representation. As a consequence of a review of this and similar findings, Nicholson (1977) decided to omit the study of syntax from his investigation of the type of

changes to text that may determine comprehension, e.g., the presence of anomalous material or the absence of a title.

On the other hand, semantic prediction at the sentence level is one of the dimensions of processing commonly thought to distinguish the good from the poor comprehenders, as studies generally demonstrate that a higher proportion of the miscues of good readers are meaningful (Malicky, 1982, p. 5-4). This fact is evident both from the earlier exploratory miscue studies (e.g., Clay, 1969), and later studies involving statistical analysis (e.g., Beebe, 1980; and Pace, 1977).

Accessing real words also loads on this dimension of meaning construction. The description of lexical access introduced in factor three is valid here, but now top-down processing plays a more prominent role in associating an appropriate response. The readers use context and a minimum of graphic information to help predict word identification which is meaningful both in terms of the reader's oral language and the author's meaning. Therefore, the readers are making sense of what they read and actively attempting to comprehend the author's message by semantic prediction. The depiction of this dimension as representing the strategies used by good readers is given further weight by research carried out by Olslavsky (1976) with grade-ten good and poor readers. The good readers used context to define a word while the poor readers simply stated their inability to define a word.

Pseudo-word accessing and a dependent variable, minor

change in the author's meaning, are inversely related to this constructive processing configuraion. In consequence, the readers who engage in semantic prediction to comprehend the author's meaning are not inclined to produce nonmeaningful responses to words, nor to make uncorrected errors that cause minor meaning changes. It must be remembered that minor changes are defined as: errors which cause changes only in sentence, not passage, level meaning, and do not involve a key content word. Hence, minor meaning change can include non-meaningful responses. Together, these two variables signal processing that is probably more concerned with decoding than with semantic association.

Finally, this entire processing dimension may be said to resemble some of the findings of a literature review study carried out by Golinkoff (1976). She reviewed research in reading comprehension to discover how good readers differed from poor readers, and concluded that they differed mainly on two variables: decoding and text organization (defined as getting meaning from larger units of text). Since this review was carried out at a time when discourse processing was still in its infancy, "larger units of text" referred chiefly to sentence units. As the proficient reader interacted with text, they strived for meaning, and further more, appeared to possess an awareness of when they understood or failed to grasp the intended message, while the less proficient readers appear to engage in reading mainly as a decoding task.

To conclude, the processing strategies represented by

this factor bridge the gap between how words are identified and the construction of larger units of meaning. Malicky (1982) includes this aspect of processing with word identification, whereas the Goodmans (e.g., Goodman and Goodman, 1977) use an analysis of how miscues are contextually constrained as a window into all aspects of meaning construction. Nevertheless, there is a significant absence of any indicator of text-level semantic construction. Hence, this factor, semantic constructive processing at the sentence level, represents a dimension of processing distinct from passage-level semantic processing.

Factor Five: Constructing-Reproducing

This dimension of processing represents how the fourth grade children processed units of text information in an exact manner. More specifically, the loadings on this factor indicate the construction of the exact meaning of text sentences (clauses) and the reproduction of those units in a verbatim fashion. The three significant loadings on this factor are presented on Table 11.

Table 11

Factor Five: Loadings greater than .4 from Oblique Rotation (PROMAX)

Text Exact Oral	.784
Text Exact Silent	.674
Text Erroneous: Associating Oral (E1)	-.446

The significant loadings on this factor indicate that processing units of information in an exact manner is a distinct

facet of how the readers processed discourse and is inversely related to recalling inexactly associated units in the recall of oral reading. More specifically the loadings on this factor indicate one way the readers abstracted the meaning of small units of text and committed them to memory and then retreived that information during recall. When engaging in this type of processing the readers attended carefully to print, associated it with the exact phonological and meaning association, encoded that information in memory and recalled it without further transformations. Although these abstracted units of text may eventually become part of semantic memory, for the moment (that is, between the time of reading and recall) they remain part of episodic memory, that is memory for that specific unit of text (see Shoben, 1980, pp. 311-312).

However, since no definitive theory of how individual sentences are comprehended exists, it is difficult to make substantiated inferences about the mechanisms of verbatim processing. And even if such a theory existed an "interface" would then be "needed between the higher level processes and the fundamental processes involved in sentence comprehension. The question is not whether these higher order processes have an effect on sentence comprehension but how they influence more basic reading processes" (Shoben, 1980, pp. 327-328). With these theoretical limitations in mind some tentative explanations will be offered of how the readers first constructed, then reproduced exact text units. It appears as if the verbatim recall in this study reflects more than a default type of lower-level processing or memorization; it is

rather an integral part of discourse comprehension influenced by top-down processing and contextual effects. The rationale for this speculation is threefold: (1) research on the type of text information proficient and less proficient readers recall, (2) the overall configuration of text information recalled by the readers, and (3) characteristics of the text units that tended to be recalled in a verbatim fashion in this particular study.

Until recently practitioners in the field believed that poor readers recalled more verbatim information than good readers. Recalls containing a lot of these units supposedly were taken as a indication of lack of integration of the incoming units into a broader semantic framework, thus reflecting lack of passage level synthesis. For instance, one of the hypotheses investigated in Beebe's (1981) study reflected this belief: the greater the amount of text information exactly recalled "the lower the comprehension score" (p. 76). But this hypothesis was rejected because "the effect of abstraction of verbatim information had no effect upon reading comprehension when taking into consideration the degree to which the reader summarized, synthesized or otherwise integrated other abstracted information with this background" (p. 188). This finding was also corroborated by Brake (1981) in her investigation of the comprehension of good and poor second grade readers as they read passages orally and silently. Furthermore, Tierney et al. (1978) found that both proficient and less proficient readers engaged in abstractive and constructive discourse processing operations when reading and recalling an

expository style passage.

Problems arise when making any direct comparisons between the present and previous research because of definitions of verbatim recall. Both Beebe (1981) and Brake's (1981) definitions include word-exact reproduced units and transformations of the text specific information. Tierney et al's. (1978) definition of abstracted information is even broader: "Abstractive processing involves selective processing, which seem prompted by the reader's attempts to 1) glean what might be considered relevant units from the text and 2) summarize the ideas in a manageable form in accordance with what can be handled by the memory system" (p. 552). In fact, abstracted or verbatim information includes all recalled units whether synthesized or not that can be traced to explicitly stated text information.

Despite these broader definitions of what constitutes verbatim recall, it is safe to assume that there is sufficient overlap between the definitions to warrant some comparisons. All three studies show that both good and poor readers engage in reproductive and reconstructive or integrative processes. Hence, it may be assumed that the readers in the present study also engaged in higher-level organizational processing which influenced their reproduction of verbatim text units. Moreover, reproduction is probably an inevitable component of certain types of discourse processing and is also determined in part by the nature of the text itself.

Exact or verbatim recall tended to be associated with

the abstraction of particular units of discourse, generally occurring on either of two occasions: recall of short colloquial clauses and partial clause recall. The readers who reproduced one or two colloquial type clauses did so apparently because they were aware of the more dramatic quality of these pithy statements within a larger text framework, not because they were attempting to reproduce a verbatim rendition of the surface features of the whole text. Examples of these units are: "At last the big day arrived", and direct speech, e.g., "The game is over" and "There it is". So at the time of recall these salient units were generated during the sequence of recall. For instance, during the recall of Soccer a child recounts that Paul found the ball and then says "something" to the team. In other words, these bits of exactly stored text, are retrieved when cued by gaps in the framework of the story sequence.

The processing involved in the inclusion of partial units of exactly reproduced text is close to certain facets of Tierney et al's (1978) definition of abstractive processing, where the reader "gleans ... relevant units from the text" and summarized these units in an appropriate manner for retrieval (p. 552). Examples from the protocols will help to illustrate what is meant by incomplete recall. The following excerpts are from the recall of a subject who read the more difficult passages and produced longer than average recalls indicating emphasis on comprehension rather than on memorization:

(1) Text: Frank, Mary and their friends are interested in track and field.

Recall: Frank and friends are interested in track and field.

(2) Text: Unfortunately for him, he landed in the sand crying with pain.

Recall: He landed in the sand crying.

This type of incomplete recall fits into the ongoing pattern of children's recall and seems to occur when the text language is similar to the child's syntax and vocabulary. In other words, the reproduction of these types of units appears to be incidental to making the unaided recall.

Factor Six: Reconstructing-Synthesizing

This, the sixth factor is a latent processing dimension which reflects how the readers synthesized larger units of information and then produced reconstructions of the text's meaning during recall. "Reconstructive memory processes" refer to rearrangements of stored text information which may proceed after the reading has ended; recall is then based on the present state of the organizational pattern in the reader's mind (De Beaugrande, 1981, p. 264). In the case of the synthesizing represented on this factor, some summaries or integrations of textual information may be made just prior to the recall and moreover may have been prompted by the necessity of producing an oral recall. The variables loading on factor six are presented on Table 12.

Table 12

Factor Six: Loadings greater than .4 Oblique Rotation (PROMAX)

Text Entailed Oral (C)	.753
Text Entailed Silent (C)	.563
Text Specific Silent (B)	-.650
Text Specific Oral (B)	-.504

The two variables with significant positive loadings on this factor are indicators of across-sentence and across-clause level syntheses, whereas the two variables loading in an inverse direction show how the readers constructed the meaning of individual text units during reading, and then transformed the meaning of these specific units by the use of synonyms, etc. When the fourth grade readers in this study engaged in synthesis during and after oral and silent reading, they linked and integrated the meaning of clauses from across the text. Integrative processing is often regarded as a feature of proficient reading (e.g., Beebe, 1981; and Brailsford, 1981).

Brailsford (1981) demonstrated that reading as active cognitive processing can be explained within the model of simultaneous and successive synthesis developed by Das et al. (1979) from Luria's (1966) theories of brain function. Describing these two categories of cognitive processing, Das et al. (1979) state:

> Simultaneous integration refers to the synthesis of separate elements into groups...The essential nature of this sort of processing is that any portion of the result is surveyable at once...Successive information processing refers to processing of information in a

> serial order. The important distinction between this type of information processing and simultaneous processing is that in successive processing the system is not totally surveyable at any point in time. Rather, a system of cues consecutively activates the components (p. 50).

Based on the above distinction, it seems as if the type of text integration that is reflected in this processing dimension may be aptly characterized as simultaneous synthesis, where the readers grasped the essential elements of the story as a single causal unit. Protocols produced by two different subjects have been selected to illustrate this facet of discourse processing and are presented on Table 13 (1 and 2).

An investigation of the protocols of the 102 children made it clear that the tendency to synthesize or to produce text entailed units (C) and to eschew the recall of specifically associated units (B) was demonstrated in two distinct forms; namely, the tendency to give: (1) a short summary type recall by retrieving few details, or (2) a longer recall integrated by chunking adjacent units of text. In the first, the summary type of recall, the reader integrated information from widely separated units in the passage producing mainly text entailed units (see Table 13-1). In contrast, the reader who produced the longer integrated protocol in 13-2 synthesized by fusing adjacent clauses and sentences e.g., He went high, high over the bar and landed crying. However, this reader also recalled a small proportion of specifically associated units; in the present example, four specific (B) and eight entailed (C) or synthesized

Table 13

Protocols to illustrate the reconstructing-synthesizing processing dimension

.1

Frank, Mary and[C] friends wanted to join the track and field// and Frank[C] wanted to join the one hundred meter dash and Mary the jumping// (then Mary) then Mary[C] won a prize// and their fr[B]iends didn't win any// just before the one hund[C]red metre dash Frank sprained his ankle// and (Frank no) Mary and her friends [C]brought him to the first aid room//

.2

Frank and Mary were [A] interested in track and field// and they wanted to do the competitions for running and jumping[C] in the next track and field competitions// because they[D1] wanted to compete against the other players// and they all had specialities [C] like the best things// so Frank was interested[B] in the one hundred metre dash// and Mary was interested[C] in the high jumping// (well) Mary and Frank and their [C] friends practiced day after day week after week for the big day// then on the big day Mary[C] was the first to compete// but she only[B] came third// and the fri[B]ends got nothing// then Frank tried[C] jumping for fun// He went high, [C]high over the bar and landed crying// because [D1]his ankle was badly sprained// Mary and her [C]friends came running and helped him to the first aid// while the one hundred[B] metre dash was being announced// (and that's all).

units.

The type of processing phenomenon that is apparent on this factor to a certain extent conflicts with the findings of two recent studies. When investigating the relationship between the discourse processing strategies of grade four readers and their comprehension score on a standardized test, Beebe (1981) found that: "the amount of information a child was able to infer, summarize or reconstruct as he read had a strong positive effect on his ability to comprehend during reading" (p. 138). Tierney et al. (1978) also showed that the more proficient grade three readers "process and generate more causal and conditional relationships which add to the coherency and organization of their recalls" (p. 566). But no such relationship between reading proficiency and the production of integrated information in recall is evident in the present research; measures of reading proficiency are uncorrelated with text synthesis, moreover, none of the dependent variables are associated with the synthesizing dimension. A note of caution, however, is warranted when comparing the present results with others as other studies employed different units of text analysis to arrive at their conclusions. For instance, Beebe's (1981) text-entailed (synthesis) category included both synthesized and inferred-instantiated material. Thus, the conflicting results may be in part due to the use of different, but nonetheless overlapping, categories and to the greater specificity of the present analyses which attempts to isolate the nature of reading process into more distinct variables or measures.

The second point of contrast with the Beebe study concerns the relationship between text specific and text entailed information. Beebe (1981) discovered that a high positive correlation existed between these two aspects of text processing and that, in fact, these aspects represented a single factor. Hence, those children who recalled more text entailed or integrated units also recalled more specific or verbatim or transformed units. Yet on the present synthesizing dimension, these processing indicators occupy opposite poles of a processing complex thereby indicating that the readers who displayed a proclivity to retrieve specific text units were not at the same time inclined to produce more integrated or synthesized units. Length of the stories read and recalled in the studies may help explain these differences; in the Beebe study the chilren read one long story -- whereas in the present study four much shorter stories were read. Hence, in the former study there was probably a greater opportunity for the subjects to recall both synthesized units and descriptive or action details; in the present study, recall of both types of information may have appeared redundant in one short recall.

Finally, the topic of recall directions and the children's interpretation of those directions must be raised. Whereas in the present study the children were given only general recall directions, Beebe (1981) gave both general and specific directions; she asked the children to recall what the story was about in their own words but also to recall "everything" they could

"remember". The latter directions plus the longer story may have precipitated the recall of both text entailed and specific information. The present more open recall directions may have allowed for different interpretations; thus, if the children recalled the main events of a story in a synthesized form they may not have considered it necessary to retrieve all that they remembered.

Factor Seven: Constructing-Transforming

The seventh factor represents a latent semantic processing dimension of constructing and then transforming the meaning of specific text units. When the children engaged in this type of processing, they probably tried to grasp the meaning of individual text units without necessarily integrating each into a cohesive framework or schema representing the entire story. Moreover, this characterization is further substantiated by an investigation of the negative loadings on the factor. Inversely related to constructing and transforming specific semantic units is a processing configuration where the readers integrated the individual textual units and therefore produced inferences, instantiations and elaborations on the explicit textual information. The variables with significant loadings on this processing dimension are presented on Table 14.

Table 14

Factor Seven: Loadings greater than .4 Oblique Rotation (PROMAX)

Partial Syntactic Acceptability	.562
Text Specific Silent (B)	.462
Text Specific Oral (B)	.423
Omissions	.400
Text Experential: Elaborating Silent (D2)	-.700
Text Experential: Inferencing Oral (D1)	-.668
Text Experential: Inferencing Silent (D1)	-.470

The highest positive loading on this factor, syntactic acceptability at the partial sentence level indicates that the readers were not predicting syntax and possibly semantics at the sentence level. And as a consequence of this piecemeal processing these readers were unable to build a stable framework to incorporate and synthesize larger units of meaning for memory so at the time of recall produced small associated units of story information which were not fully integrated or synthesized. Moreover, the appearance of omissions on this processing dimension yields further indication of the mechanisms of the processing strategy. Hence, the nature of these omissions will be considered below.

Goodman and Gollasch (1980) demonstrate that omissions cannot be viewed as a single miscue category but are "complex manifestations of the reading process" (p. 28). These authors identify two general categories of omissions which may occur during oral reading: deliberate omissions are characterized as being the reader's way of coping with the problems of word identification. Thus, "if readers are producing deliberate omissions,

they are... choosing among alternatives" (Goodman and Gollasch, 1980, p. 16) -- one possible alternative being the production of pseudo words. In contrast to deliberate errors, non-deliberate errors tend to occur when the readers can easily master the reading material, and are not experiencing any kind of processing overload and are, therefore, predicting both syntax and semantics. Hence, these omissions are typically of "easy" or "common" words usually of optional elements such as articles.

But neither of the two types of omissions depicted above accurately characterize the omissions which form part of this factor. An analysis of the protocols showed that only two children in the sample of 102 seemed to have engaged in omitting less familiar words as a deliberate strategy. Furthermore, the type of omissions reflected on this dimension cannot be non-deliberate fluent oral reading errors as this variable, omissions, loaded on the same dimension as the indicator of partial syntactic acceptability which is not a characteristic of fluent readers. Moreover, an examination of the correlation matrix shows that the omission category correlated +.400 with both partial prediction (syntactic and semantic) and partial sentence synthesis but is uncorrelated with sentence level semantic or syntactic processing, again demonstrating that at least some of these omissions are not a correlate of independent reading. The type of omissions indicated on this dimension are classified as "non-deliberate frustration omissions" and were typically made when the children experienced difficulty with word identification

and possibly also with the conceptual load or content.

The grade four readers engaged in the processing strategies reflected on this semantic constructing-transforming dimension as they neared frustration level to cope with the assessment task; although the children may have known that the text was too difficult for them, they attempted to comply with the request to read to the end and then give a recall. Thus, the subjects found it necessary to predict, associate, and synthesize small semantic units because they were unable to cope with larger units. As many of the elements of the text sentences were unpredictable, words and lines of text were omitted and were not monitored although they resulted in sentences that were not syntactically acceptable. As a consequence of this type of ad hoc processing during reading, the readers never managed to develop an exact causal schema for the written narrative with the result that recalls were produced consisting of specific units of text and few inferences and elaborations.

An example of the oral reading errors and recall of a subject are presented on Tables 15 and 16 to illustrate the characterization of the processing strategies developed above. The child began to engage in the strategies reflected by this dimension from the point in the story where the first line is omitted. The omission of this line then triggered the subsequent omission of "him" as the reader attempted to make sense of the cojoined sentence fragments. An almost identical processing sequence may account for the omission of a second line

Table 15

Oral reading errors which illustrate the type of strategies reflected on factor seven

Form C

Cross-Country Skiing

were
Tom and his sister Sonia are really keen on cross-country
skiiing. Last winter Uncle Ben promised to take them on a
real wilderness adventure when they could ski six miles.
Every weekend they practiced skiing. By the end of February
they could ski for three hours without becoming exhausted.
Then Uncle Ben told them about his winter-camping surprise.
the
He had a big winter tent hidden away in a sheltered spot in
off
the bush, four miles from the nearest road. Skiing was the
only way in there. The children were thrilled and diligently
began to prepare their skis, warm clothes and pack sacks.
One morning, Uncle Ben collected them in his ramshackle old
truck. On the way he told them about his tent with its wood
stove and supplies of food. When the road ended, Uncle skied
first and broke trail through the trees. His niece and nephew
senses
skied after him admiring the winter scenes. Sonia imagined
the fairland
she was in a glorious fairyland. The frozen forest was glis-
tening in the winter sunshine. Tom was so busy looking about
hisc
him that he collided with a tree and broke a ski tip. Uncle
was prepared for such accidents and had brought along a spare
ing
tip so they lost no time and arrived at the tent well before
nightfail
nightfall. Soon they got the stove burning brightly and ate
supper. That first night sleeping in a tent was especially
associational
exciting. The silence was broken only by the occasional
howling of a distant wolf.

Table 16

Recall of Cross-Country Skiing which reflects the type of strategies evident on the seventh factor

Tom and Sonia loved [B] skiing cross-country// and they prac[B]ticed skiing// and at the end of February they could [B] keep going for hours without getting exhausted// and then [B] they went// (and then and) they [B] lit the stove// and they slept in [B] the tent// and it was so exc[B]iting// (and then and then they they then the uncle then the uncle Ben) uncle Ben (uh) his (a him uh) ram-shackle [A] truck// and he [B] went// (uh) and then [B] he skiied (and it all uh)// and then he came to [B] the tent// (and they and that's all)

followed by the omission of "and". Furthermore, these omissions resulted in sentence segments that were only partially syntactically acceptable.

This child's recall of Cross-Country Skiing reflects the processes engaged in during the passage reading. The middle of the story is omitted; probably because during the reading of this portion of the story, the child engaged in partial syntactic prediction and omitted words and lines. Nevertheless, this child recalled the end relating how Tom and Sonia slept in the tent. However, he then attempted to recall some details from the development but because of his failure to conjoin these elements the recall is fragmented.

While it cannot be said that the processing strategy exemplified by this reader is devoid of any type of synthesis because these units reflect some level of organization since they are not randomly sequenced, the question is then to specify the level or type of synthesis.

On the basis of Brailford's (1981) interpretation of Das et al's. (1979) concepts of simultaneous and successive synthesis, it seems plausible to depict the synthesis of information reflected by the seventh factor as successive. When the readers engaged in these processing strategies, they probably never grasped all the main elements of the story as a single causal unit -- if for no other reason than omission of important elements during reading. For instance, although the reader selected to exemplify this constructing-transforming processing strategy,

recalled the beginning and end of the story in detail, his recall did not demonstrate that he had integrated all of the units into a causal store schema. For instance, he failed to link all the recalled elements, or to produce summary statements or inferences.

Factor Eight: Error Elaborating

The eighth factor reflects how the readers engaged in erroneous processing and elaborating. As the readers read the passages orally they made errors which resulted in major changes to the author's meaning which interfered with their ability to construct the gist meaning of the entire passage but nevertheless, they managed to associate the overall topic of the passage, e.g., horses or skiing. Then when asked to give a recall, those readers who lacked a coherent understanding of the story reconstructed a recall by combining the bits and pieces of abstracted information and elaborated it on the basis of their background knowledge. In other words they took liberties with the textual information. The variables loading on this factor are given on Table 17.

Table 17

Factor Eight: Factor loadings greater than .4 from Oblique Rotation (PROMAX)

Text Erroneous: Inferencing-Synthesizing Oral (E2)	.879
Text Experential: Elaborating Oral (D2)	.709
Major Meaning Change	.576
Passage Level Semantic Acceptability	-.461
Minor Meaning Change	-.448

The variables with negative loadings on this factor -- miscues that are semantically acceptable at a passage level, and miscues which cause minor changes to the author's meaning -- show that constructing meaning and comprehending at a text level is inversely related to erroneous processing. This finding is confirmed moreover, because one of the dependent measures of reading comprehension, uncorrected oral reading errors which cause major changes to the author's intended meaning, is part of the same underlying dimension as vaguely related or erroneous material in recall. But it should be emphasized that while the indicators of meaning change, elaborating, and erroneous synthesizing and inferencing signal that the children were failing to understand the intended message, they do not indicate that the children were unable to reconstruct some meaning from the text. Hence, there is a mismatch between the meaning the author intended to convey and the message as understood by the reader.

Comprehension and meaning are also components of the analytical framework of a study carried out by Goodman and Burke (1973). These authors carried out a study of oral reading mis-

cues with ninety-four low, average, and good readers from five grades: two, four, six, eight, and ten. Overall they discovered that "Miscues which were semantically acceptable usually altered only slightly the meaning of the text" (p. 149). This finding corroborates the loadings on the present factor pointing to the codetermination of the ability to derive meaning from the text and to understand the author's meaning by a single dimension.

The evidence supplied by Beebe (1981) on the relationship between erroneous and elaborated information recalled following oral reading and the comprehension score on a standardized test also has some bearing on this error-elaborating dimension: "the amount of information that was vague and unclear (elaboration) did not seem to matter relative to verbatim, integrated, and erroneous information" (p. 139) but on the other hand, "the amount of information erroneously interpreted while reading had a definite negative effect on ability to comprehend" (p. 138). Although in the present study, the inclusion of errors and elaborations in recall following oral reading is related to certain measures of comprehension (meaning change), it is unrelated to comprehension on a standardized test (GMR). Yet once more, it must be remembered that what Beebe classified as errors and elaborations differs in some respects from the present classification.

Furthermore, the processing phenomena of the present dimension is close to one of the reading situations depicted by Giboney (1979) in his theoretical study of the encounter between

reader and text. In Giboney's terms, the strategies of the readers in the present study might be explained by describing the readers as having "contextualized the text...in terms of a type of episodic encounter that was inappropriate for the text" (p. 226) which then lead to miscommunication. Giboney (1979) asserts, furthermore that problems of associating the text and integrating it with background knowledge are apt to arise when the reader and the writer bring "a different cultural/institutional/ personal meaning-context to their respective sides of the page" (p. 218). In the reading situation experienced by the children in the present study, certain aspects influencing the extent to which common ground had to be negotiated were controlled; for instance, an effort was made to match both the content and the vocabulary of the stories to the knowledge and experience of the children. Moreover, this is substantiated, since these children did not have problems integrating the text associations with background knowledge as evidence by elaborations and meaning reconstruction rather, their problem was one of reconstructing information that was adequately constrained by the supplied information. The roots of these readers' failure to comprehend may possibly be found in inefficient meaning construction strategies such as failure to predict the passage meaning.

In addition to mutual knowledge between reader and writer, the reader's intention is also crucial in directing the course of the reader-text encounter. According to Giboney (1979) the reader may choose to read the text to comprehend the writers

intended message or he may choose to ignore it and read for his own purposes or he may equally well choose to terminate his reading if for any reason he has problems contextualizing the message contained in the print. At present, it is assumed that when engaging in this reading task the grade four readers accepted the directions to read and recall the story, and so intended to get the appropriate text message; however, unlike the situations described by Giboney these readers probably did not feel free to terminate the encounter when they experienced problems understanding the message conveyed by the text. Hence, the present complex of reading processes may be manifested whenever readers are compelled to read material which is too difficult for them.

An example will help to clarify the kind of processing complex reflected on this, the eighth factor. A record of the child's oral reading and recall of the level four story, The Horse, is presented on Table 18. During the oral reading the child made a total of seven uncorrected errors, four of which caused major meaning changes (marked with an asterisk on Table 18) and none of which caused minor or sentence level meaning changes. For approximately the first half of the story the reader is engaging in effective strategies and is monitoring meaning and comprehension. The problems of comprehension began with the omission of the line in the text commencing with "middle...". This omission disrupted the communication of a key idea, hence it caused a major meaning change, although the gist meaning of the story in terms of the readers oral language was preserved.

Table 18

Example of oral reading and recall which illustrates the reading strategies reflected on factor eight

The Horse

Angie and Scott live d in an apartment in the city. One evening they were chatting happily about horse riding lessons on their uncle's farm last summer. Every morning the children used to watch the horse grazing in the field and dream about riding her. Then at supper c one evening Scott asked if they could try riding

immedtately c
Uncle then planned to start teaching them to ride immediately.

and the very next morning the children dressed in their jeans and waited for their uncle. He arrived carrying a bucket of oats

c find c
to help catch the horse., When they got to the field the horse was standing right in the middle. Uncle rattled the bucket and she trotted towards the oats* and began to eat. Then uncle grabbed her nostrils and held firmly. In an instant,

bridge*
he slipped on the bridle. "May we both ride back to the farmyard?" asked Angie. Scott and Angie straddled the horse's

briddle*
back very carefully. Then their uncle held the bridle and

table *
led them back to the farmyard. At the stable, they all helped

the and
put on the saddle for their first riding lesson. They soon became experts at their new hobby.

A
Angie and Scott they wanted (to ride horses) to ride horses on their C uncle's farm// (and when supper) and when they D were having supper Angie asked uncle/ if they could ride C those horses on his farm// (so uncle um) after supper they went to feed E2 Angie's Uncles horse/ and they were C riding the horses// and they E2 went over the bridge// and Angie's E2 uncle (left) left up the bridge// and (um) Angie E2 (took) pulled him back.. and they went D1 riding horses again// and when they D1 got back to the field// they saw more D2 horses// and they D2 liked the horses very much// (that all I know)

It is the mispronunciation of "bridle" as "bridge" and then as the pseudo word "briddle" that started the disruption of both comprehension and meaning. Yet, the reader made no obvious attempts to monitor these anomalies in the story schema. This failure is perhaps a result of some form of processing overload as earlier barriers to comprehension were monitored, e.g., the miscue "find" for the field was corrected in the sentece "When they got to the field..." In contrast, this readers seems unable to monitor when the meaning of the incoming message is severely disrupted. Monitoring may then be seen to operate only when the task demands are below a certain processing threshold.

In conclusion, the readers who engaged in the erroneous and elaborating processing evident on the eighth factor were unable to construct adequate meaning or comprehension during oral reading, however they were able to reconstruct a meaningful recall. Thus, in the presence of miscues that interfered with meaning and comprehension during oral reading, meaning was more likely to be reconstructed after the event as the children endeavoured to put together and elaborate a recall.

Conclusion

Eight factors accounted for fifty-six per cent of the variance in the processing evident in the oral reading and recall of the grade four students in this study. These factors were interpreted as reflecting underlying processing dimensions or strategies which resulted in variability in reading behavior.

It will be demonstrated in the final chapter that a certain degree of construct validity has been established for the eight reading processes elaborated in the theory of assessment of reading processes.

CHAPTER VIII

CONCLUSIONS AND IMPLICATIONS

The Study in Review

Kemeny (1959) defines the scientific method as a "cycle of induction, deduction, and verification, and (an) eternal search for the improvement of theories which are only tentatively held" (p.176). Since establishing construct validity is tantamount to doing science, and since the present investigation was carried out to establish the construct validity of the theory of assessment of reading processes, it necessarily follows that this study is concerned with the scientific cycle of induction, deduction, and verification.

Induction was the main tool used to build the theory of assessment of reading processes which was outlined in Chapter 3. First, the factors which were judged to impinge on the interaction between reader and text were mapped out : prior knowledge of the reader, nature of the text, and the assessment situation involving reader-examiner interaction. Then, how the reader interacts with the text in a bid to construct and later reconstruct the written message was outlined in detail as a set of eight cognitive and linguistic processes. This was accomplished primarily by synthesizing the work of other researchers and theorists in reading (e.g., Goodman, 1976 ; Smith, 1978), and in other areas such as cognitive psychology (e.g., Ortony and Anderson, 1975). But this theoretical synthesis portraying reading as an active exchange between the information on the page

and the reader's background knowledge had at best only a tentative basis in reality. The next step was then a metatheoretical one, where methods of establishing construct validity for a theoretical construct were explored.

In the fourth chapter, it was concluded that any evidence including test construction could be offered as components of the ongoing exercise of establishing validity for a theory. Moreover, because of a paradox involved in measuring a scientific concept, the construct validation of a theory cannot be separated from the validation of the instrument which collected the validational evidence. Hence, this study could be seen as serving a dual purpose : establishing the construct validity of the theory of the assessment of reading processes and the instrument designed to measure those processes, the ARP. This goal has been accomplished in the remaining chapters (5, 6, and 7). In the fifth chapter, for instance, the ARP was constructed to measure the eight reading processes. Other aspects of the theory of reading processes, such as text structure and content, were also taken into consideration in the instrument design. Chapter 6 then presented the procedure of a factor analytic study in which the ARP was employed to collect empirical evidence from a sample of 102 grade-four readers. Because this was an exploratory study, no definite hypotheses were deduced from the theory to be tested ; however, six dependent variables were included in the analysis to act as outside criteria for the processes. The outcome of the factor analysis, consisting of eight factors, was interpreted in Chapter 7 and related to similar research findings.

It will be the primary purpose of the remainder of this chapter to pull together the threads of the validational evidence for reading processes from the factor solution, and to investigate the influence of single processes or groups of processes on reading achievement and on learning to read. But in addition, since the reading theory and assessment instrument are only tentative steps along the path to scientific knowledge, suggestions for their modification in future research will be made. Finally, some recommendations for educational practice will be made on the basis of the present work.

Construct Validity of Reading Processes

A certain measure of construct validity has been established for each of the postulated reading processes through the investigation and interpretation of the eight underlying dimensions or factors. As each process is discussed below, two points will be investigated : (1) the source of the validational information in terms of the factors, and (2) the co-occurrence of particular processes. This discussion will be followed by a discussion of the part played by the dependent variables.

Attending

Information related to four aspects of attending to graphic information is represented in the data. If a child attended to a high proportion of the graphics, he was also likely to do so systematically and attend to the graphics in the initial, medial, and final positions in the word. This attending strategy is supported by

four loadings on Factor Two : Grapho-Phonic Processing -- attending to total graphemes (.756), and attending to graphemes from the initial (.706), medial (.652), and final (.539) word parts.

This attentional behavior, however, is closely linked to the ability to associate correct phonemes to the graphemes attended to, and to the ability to associate appropriate units of sound to initial, medial, and final word parts.

Analysing

Attending is closely related to the process of analysing. In order to attend productively to the print, the reader has to analyse print into various units. The strongest evidence in favor of this attending-analysing partnership comes from the information on Factor Two. As indicated above, the readers who attended to the highest number of graphics also analysed the word into parts, attending systematically to the graphics in the initial, medial, and final positions. Furthermore, this analysing appears to have been effective for the purposes of sound association, as these readers were also more successful in associating the expected sound units to the grapheme units in the word. Attending behavior for the initial, medial, or final part was scored if any letter in that section was attended to ; however, to get the equivalent score for sound-symbol association, the whole of the initial, medial, or final parts had to be pronounced exactly the same as in the stimulus word. For example, the child who read punestured for punctured got credit for attending to the three parts of punctured, but got only a score for two word parts -- initial and final -- on sound-symbol association.

There is also some evidence that the more proficient readers

(as determined by achievement score, level of passages read, and self-correction) analyse words into parts. Accessing partial words (.460) loads on Factor One : General Achievement-Monitoring ; when this loading was interpreted, it was argued that analysing words to enable partial phonological access was typical of the proficient reader's behavior when longer, less familiar words were involved.

Hence, the greater part of the evidence shows that attending-analysing behavior is not strongly related to meaning or comprehension ; no indicator of semantic processing or comprehension is in evidence on the second factor on which the eight indicators of attending-analysing load. However, one indicator, accessing words in parts, loads on the first and third factors as parts of underlying processing dimensions which include semantic construction. Factor one reflects a latent dimension where the more proficient readers read in an exact manner and monitored themselves each step along the way. And on Factor Three : Unit Accessing-Sequencing, partial accessing loads on the same dimension as accessing phrases, mis-sequencing, and inserting words ; this apparently indicates the grade-four readers' tendency to alternate strategies between closer attention to the graphic and semantic/syntactic prediction. The fact that analysing was related to semantic processing even for a small proportion of word processing may show that these grade-four readers were utilizing semantic as well as graphic information to analyse words.

Associating

Two distinct kinds of association are measured by the ARP : sound association and meaning association. Judging from the evidence

gathered in the present study, it appears that these two branches of association are largely unrelated. Sound association in word identification is strongly in evidence in Factor Two : Grapho-Phonic Processing, appearing along with attending-analysing behavior ; whereas associating meaning to word units appears on Factor Four : Semantic Constructive Processing Sentence Unit. Thus, producing miscues that are real words (lexical items) rather than pseudo-words is the main evidence of semantic association in word identification.

Attempting to utter words that are meaningful is related to both comprehension and meaning, as the information on the loadings on Factor Four indicates :

(1) Accessing real words (.852) loads positively with errors that are semantically acceptable at the sentence (.815) and partial sentence (.876) levels, thus signifying that associating meaning to word units is likely to occur when the readers use their prior knowledge to interpret the graphic symbols in terms of their own cognitive schemas. In addition, sentence (.793) and partial sentence (.883) level prediction also forms part of the same dimension as word association, once more highlighting the use the readers made of their prior knowledge to project their meaning onto the page.

(2) However, the loading of errors that do not change the author's meaning (.708) on this factor demonstrates that semantic word association is linked to the readers' efforts to comprehend the intended message in the text. This clustering of processing indicators points to the fact that semantic association at the word level is a key process linking graphic input with semantic constructive processing

of larger units such as sentences. These loadings, furthermore, present evidence for top-down or conceptually driven word processing through the use of the sentence context.

Two other types of semantic association of sentence or clause units are in evidence on Factor Five : Constructing-Reproducing, and Factor Seven : Constructing-Transforming. In the first type, associating text in an exact manner, the readers abstracted the meaning of clausal units verbatim, and reproduced them in exact form in their recalls. This is clearly a distinct dimension of meaning association in both oral and silent reading, as the indicators of associative-reproductive processing load on their own on the fifth factor. Furthermore, the exact strategy is inversely related to the tendency to reproduce erroneous association of clausal units following oral reading.

Support for the second kind of association of clausal units is found on the seventh factor, because text-entailed (B) units recalled after both oral and silent reading load positively on the constructing-transforming dimension. When information was recalled within these text-specific categories, it was hypothesized that during reading the readers abstracted the gist meaning of clauses and associated it with knowledge in their semantic memory, then recalled that clause in a transformed manner, rearranging or substituting lexical items while conveying a synonymous meaning. This transformation of clausal meanings is part of the same dimension as partial sentence-level syntactic prediction and omissions, and is inversely related to inferencing in oral and silent reading -- thus suggesting

that on this dimension, at least, transformation of individual clausal units is not conducive to the integration and instantiation of the passage meaning using background knowledge. This recall phenomenon is further substantiated because on Factor Six : Reconstructing-Synthesizing, both oral and silent text-specific categories (B) load negatively with respect to variables indicating synthesizing meaning from across the passage.

Sequencing

In this study, sequencing refers to the perception and association of visual information in the exact order in which it appeared in print, and includes both word and letter order. From the loadings on Factor Three : Unit Accessing, it seems as if the tendency to engage in correct sequencing forms part of a processing dimension which also includes the tendency to focus on words and to access them immediately as words or pseudo-words, and is inversely related to mis-sequencing and the tendency to access partial words or phrases.

Because both words and letters were categorized as a unit, it is not possible to judge precisely what proportion of incorrect sequencing is due to the prediction of the semantics and syntax of phrase units. Nevertheless, it is probable that many of the occurrences of mis-sequencing are determined by the prediction of the child's oral language patterns because : (1) the accessing of phrases and mis-sequencing are correlated (.505), and (2) Clay (1979) says that in her analysis of the oral reading of third-grade students, "reversal of letter order was rare" (p. 154).

Predicting

According to many advocates of psycholinguistic theory,

prediction is the most important process involved in constructing the author's meaning. The two indicators of prediction, sentence and partial sentence level prediction, loaded on Factor Four : Semantic Constructive Processing-Sentence Unit. The other processing indicators in evidence on that dimension show clearly that predicting is apt to occur in the presence of meaning synthesis at the sentence (.815) and partial sentence (.876) levels, along with the semantic association of words. In addition, prediction is related to the presence of miscues which cause no change in the author's meaning (.788). The composition of this processing dimension points to a relationship between the ability to comprehend the author's message and the ability to predict on the basis of both knowledge of syntax and semantics. The tendency to predict on the basis of syntax alone plays a negligible role in determining prediction, and did not appear as part of the fourth factor.

At present, none of the measures of processing or of meaning change are used as indicators of passage level prediction. Consequently, this latter level of prediction was not accessed.

Monitoring

The readers who play the "psycholinguistic guessing game" (Goodman, 1976) first predict or anticipate the printed message, but also constantly monitor these predictions in terms of the semantic framework they are constructing for the passage content. Hence monitoring, manifested through self-corrected miscues, might have been hypothesized to load with prediction. In the present analysis, however, this often postulated predicting-monitoring partnership did not emerge as part of a single reading strategy.

Monitoring loaded (.460) on Factor One : General Achievement-Monitoring, thus indicating that the more proficient readers engaged in more self-correction. This finding is in agreement with the results of several miscue studies (e.g., Goodman and Burke, 1973) and with more recent work in meta-cognition (e.g., Brown, 1981), and is of particular interest as it may form part of what Clay (1979) calls "a self-improvement system" whereby a child may teach himself to read. Furthermore, this capacity to learn from feedback by monitoring one's ongoing reading is probably part of a more broadly based meta-cognitive dimension which may be related to a particular child's level of learning and understanding.

This awareness of cognitive functioning is intimately linked to the ability of learning to learn. This topic was of concern to Bateson (1972) for many years, as he attempted to find the roots of organic learning, which he saw as a trial-and-error reality fitting process (p.284ff). This process he first conceptualized as deutero-learning, where an individual has to acquire the habit of looking for certain contexts rather than others, and of "punctuating the stream of events to give repetitions of a certain type of meaningful sequence" (p. 166). Later in life, Bateson wished to bypass deutero-learning in favor of trito-learning, operating at a third level of abstraction from concrete behavior and defined as "learning to learn to receive symbols" (p. 249). In reading, part of learning to receive symbols is no doubt monitoring or self-correction. Bransford (1979), for instance, asserts that learning to learn "is extremely important because it suggests that efficient learning and high intelligence might involve sophisticated skills that some people have developed and

others can develop" (p. 237) -- thus suggesting that processes such as monitoring might be acquired through mediated learning or tutoring.

There is, at present, one major weakness in measuring monitoring : many of the errors in oral reading may be corrected internally as a child may not wish to interrupt the flow of reading by regressing to make an audible correction. In other words, miscue analysis does not account for accommodation or correction of errors which may happen after the reading of a particular line of print.

Synthesizing

In the present work, synthesizing refers to both phonological and semantic synthesis. Sound synthesis, i.e., accessing words as phonological units, loads on Factor Three : Unit Accessing, and forms part of the same dimension as perceiving letters and words in exactly the same sequence as in print. This co-occurrence was accepted as a reflection of an "interpretative" strategy of immediate word identification (Massaro, 1975).

Semantic synthesis of sentence (.815) and partial sentence (.876) units load on Factor Four : Semantic Constructive Processing Sentence Level, along with other indicators of meaning construction and comprehension at the sentence and word levels. Synthesis at the passage level is part of another processing dimension. In fact, it is inversely related to erroneous processing and elaborating on the eighth factor.

Reconstructive synthesis at the across-sentence level in oral and silent reading loads on Factor Six : Reconstructing Synthesizing. The text-entailed category is taken as the indicator of synthesis in recalls, and accounts for a very wide range of synthesis -- from a

chunking of information from two adjacent clauses to a statement of the topic or kernel of the story. It is of interest that this facet of discourse processing is isolated on one factor and is inversely related to the production of text-specific units. This is contrary to Beebe's (1981) findings, where the recall of entailed or synthesized units was generally accompanied by the recall of details or text-specific units. This recall phenomenon was not manifest in the present study, partly because the stories read were much shorter than in Beebe's study ; hence the children either recalled a short summary recall involving indicators of synthesis or a longer, more detailed account -- but not an integration of both.

In summary, three types of synthesis were identified in this validational study : sound synthesis, sentence and partial sentence synthesis, and passage level synthesis measured in two ways : (1) semantic acceptability of errors at the passage level, and (2) text-entailed information in recall.

Inferencing

Two indicators of inferencing were considered in this study : the production of information in the text experiential category (D_1) in the recalls of oral reading, and its counterpart in silent reading. Information in these recall categories (D_1) "involves the generation of new propositional knowledge from a network of propositions which is given -- from prior discourse, from context, or from stored knowledge of the world" (Fredericksen, 1977, p. 68). Therefore, inferencing is a reconstructive process where the reader adds to the information gleaned from the page.

The indicators of inferencing-instantiating behavior are in

evidence on Factor Seven : Constructing Transforming, where inferencing in oral reading (.668) and inferencing in silent reading (.470) form part of the same processing dimension and are inversely related to processing smaller units of text meaning in a transformed manner. Inferencing in both modes of reading appears as a relatively distinct processing complex, although elaborating in silent reading loads on the same factor ; this may suggest -- for these fourth-grade readers at any rate -- that the addition of information to text in silent reading, whether textually constrained or not, is part of the same processing dimension.

Independent and Dependent Variables

When presenting the design of the factorial validity study in Chapter 5, six of the sixty variables were designated as dependent variables ; the remaining fifty-four were considered as independent variables. The choice of which variables to designate as either dependent or independent in scientific research depends on the conceptual framework of the researcher, which involves some notions of a sequential or causal chain and in the final analysis rests on an arbitrary decision. Two examples will help to clarify this issue :

(1) Children with poor self-concepts are often those who experience problems in learning to read. Either variable may be seen as the causal factor or independent variable. For instance, if reading is seen as the dependent variable, it is hypothesized that failure to make progress in reading is a result of a negative self-image.

(2) In sociolinguistic studies, multilingualism may play the role of dependent or independent variable according to the focus of the research : as a dependent variable it may be characterized as being

the consequence of age, schooling, etc ; and as an independent variable it may be depicted as both causing and reflecting behavioral and cultural discontinuities.

In educational and psychological measurement, dependent variables also act as outside criteria to judge the adequacy of a new measure or its relationship to conventional measures of that trait. In the present validational study, the dependent variables were the measures of reading success which would then be determined by the reading processes. These dependent measures were : the comprehension score on the Gates-MacGinitie Reading Test (GMR) ; the total proportion of miscues in oral reading ; the level of passages read ; and no change, minor, and major changes to the author's meaning through oral reading miscues. Their role in the present findings will be discussed below, in addition to three other topics : the parts played by erroneous processing and elaborating, IQ, and factors without dependent variables. It should be noted, however, that both dependent and independent variables formed part of the same factor analysis, and that determination or causality is inferred from factor membership.

Dependent Variables. Three of the dependent variables -- score on the GMR, level of passages read, and proportion of miscues -- load on Factor One : General Achievement Monitoring. It appears as if the standardized reading score and the level of passages read were determined by the readers' ability to read accurately and to monitor their reading each step along the way. Furthermore, this aspect of reading success is impaired by a high proportion of miscues and erroneous association and reconstruction of text information during

silent reading. In the light of an earlier discussion, it is of interest that one of the dependent variables, viz., proportion of miscues, could also be considered an independent variable.

The dependent variables which acted as outside criteria of comprehension -- no change, minor and major change to the author's meaning -- loaded on two factors. No meaning change and minor meaning change load on Factor Four : Semantic Constructive Processing Sentence Unit. Miscues which do <u>not</u> interfere with the author's meaning appear to have been determined by the readers' ability to engage in meaning construction at the sentence level or below. This underlying processing dimension consists of the ability to predict, to synthesize the meaning of sentences, and to associate word meanings. Furthermore, this sentence construction dimension is inversely related to the minor meaning change or miscues which interfere with sentence meaning, but not with the communication of the gist of the passage. Hence, minor meaning change was determined by the readers' failure to predict and synthesize sentence meaning, but also by disregarding semantics and associating pseudo-words.

Minor meaning change also loads on the eighth factor, along with the semantic acceptability of errors at the sentence level, but is inversely related to an error-elaborating factor on which major meaning change appears. Once again these relationships between factor and variable are to be expected. Minor meaning change appears to have been partly determined by the readers' ability to construct the gist meaning of the passage, whereas major meaning change was determined by their failure to maintain that gist meaning and to engage in erroneous reconstructive processing and elaborating.

Elaborations and Erroneous Discourse Processing. Some dependent variables focused on the degree of congruence between author's and reader's meaning. From earlier discussions it has been shown that the reader can construct meaning which differs from what the author intended. This aspect of text processing was measured in the study by six processing indicators, elaborating (D_2), and faulty association (E_1) and reconstruction (E_2) in oral and silent reading.

Elaborations occurred in recall when the readers tried to embellish their recalls and to reconstruct the parts of the story which were not fully understood, or which were forgotten, by utilizing information from an associated schema in semantic memory. But elaborational behavior in oral and silent reading may represent different processes, as they do not appear on the same factor, but rather on two separate factors. In the recall of oral reading, elaborating is a component of the "error-elaborating"dimension, whereas in silent reading it forms part of a reconstructing dimension which includes inferencing in both oral and silent reading.

Erroneous discourse processing in oral and silent reading also loads on separate factors. This processing is inferred from two categories of recalled information : the first refers to faulty association of units of text, e.g., recall of erroneous facts, and the second includes erroneous syntheses and inferences. These two types of erroneous processing in silent reading are inversely related to achievement and monitoring behavior. In oral reading, however, the two facets of erroneous processing do not appear as part of the same factor.

Text erroneous-association is inversely related to reproduction of exact units of text, whereas text erroneous-synthesis/inference is a component of Factor Eight : Error-Elaborating.

IQ and Reading Processes. IQ has been variously defined over the years. To the extent that it may involve cognitive processing, it may reflect the same processes as are assumed to underlie the act of reading. Hence, IQ was considered as an independent measure. More specifically, a traditional measure of intelligence, the full-scale score on the Lorge-Thorndike Intelligence Test, was included as an independent variable in the analysis to investigate the relationship between that score and the reading processes in determining reading comprehension.

The results reveal, however, that the intelligence measure only loaded with one of the eight processes : monitoring. In fact, the LTI score is strongly related to reading comprehension as measured by a standardized test. But this is possibly an artifact of how "intelligence" was measured. No doubt, if it had been assessed by a more dynamic process measure such as Feuerstein's (1979) dynamic measure (LPAD), the relationship between IQ and reading processes would have been closer.

Decoding. Some of the constructed factors did not include a dependent variable among the factor loadings. This raises the question as to what this particular factor means in terms of reading behavior. Certainly the factor is indicative of an underlying construct based on variables that were given theoretical significance. The fact that the factor did not include a dependent measure does not mean that that particular construct does not relate to reading. It is possible

that the measures of comprehension are not the most adequate. However, it is also possible that the relationship is a matter of kind rather than degree. Certainly, the dependent variables do not form an integral part of the underlying construct ; therefore the relationship is not direct. The nature of an indirect relationship will be examined in terms of Factor Two : Grapho-Phonic Processing. This particular factor is chosen because of the controversy that surrounds the relationship of grapho-phonic processing to reading achievement. This topic has been much debated by educators since the turn of the century (Smith, 1965), and questions such as, Should early instruction lay emphasis on phonics or getting meaning from text ? are still a hot issue (see Chall, 1967, for a detailed discussion).

Resnick and Weaver (1979) discuss "the centrality of decoding to early reading," and conclude that researchers who address this issue take one of three positions that vary along a continuum from :

> (a) learning to decode (i.e., using knowledge of phoneme-grapheme correspondences to recognise words) is early reading ; to (b) learning to decode (through phonics instruction or some similar approach) is important because it helps develop sensitivity to orthographic regularity (i.e., recurrent spelling patterns), which is important in the transition from early reading to skilled reading ; to (c) learning to decode is at best incidental to becoming literate, and at worst may interfere with acquiring reading skill (p. 8).

Early reading is decoding : theorists and researchers who maintain this view see learning to read as an epiphenomenon of learning

to speak. For instance, Mattingly (1979) maintains that phonological awareness is a prerequisite to success in learning to read ; and Liberman and Shankweiler (1979) argue that "to learn to read, children must map the written word to the spoken word" (p. 109). Furthermore, this mapping is accomplished on a systematic phoneme-by-phoneme manner. At all phases of reading, reading is primarily the process of decoding graphemes to phonemes, but once this is accomplished the semantic and syntactic skills of oral language take over.

This view of decoding is not borne out by the present results, at least for fourth-graders' reading. Firstly, grapho-phonic processing is not related to meaning or comprehension ; and secondly, phonological units and grapheme units larger than single phonemes or letters appear to have been employed when executing sound-symbol relationships. However, level of reading proficiency probably determines this relationship : in the first year of formal instruction graphic decoding is a necessary step to meaning (e.g., Biemiller, 1970 ; Weber, 1970) ; but after this, attention to graphics is inconsequential to reading comprehension (e.g., Beebe, 1981).

Reading and Orthographic Regularity : Even though skilled silent reading may not necessarily involve phonological encoding, it involves attention to recurring letter patterns in printed language. Therefore, rather than being aware of phonological segmentation a skilled reader must be aware of and utilize orthographic regularity. Joula et al. (1979) hold that "the route to rapid word identification and skilled reading depends on the development of visual processing skills that make use of orthographic regularities or direct recognition of frequently occurring letter clusters and words" (p. 105). Also,

Venezky and Massaro (1979) claim that identifying words in this manner is the only skill unique to reading.

This position is in part supported by the loadings on Factor Two : Grapho-Phonic Processing, as they seem to represent a processing configuration that includes matching larger letter patterns to phonological patterns. Yet, the necessity of this orthographic decoding en route to meaning is not substantiated.

Decoding is not Central to Reading : Decoding is of secondary importance, even in initial reading, towards the task of getting meaning from print. Written language is viewed as an alternative form of communication and, like learning to talk, it should be learned in a functional context. In addition, emphasis on decoding simply impedes progress in the enterprise of learning to bring meaning to print, because use of graphic information should always be subservient to higher-level processes such as predicting. The most vocal advocates of this top-down view of reading are the Goodmans (e.g., Goodman, 1980).

Partial support for this view may be derived from the findings presented, so far as systematic decoding of the graphics appears to bear little relationship to meaning. On the other hand, in factor one, part word accessing was related to proficiency in comprehension. But two points related to the nature of the task must be remembered : (1) the higher achievers engaged in partial word accessing on longer, less familiar words that presumably would be difficult to predict ; and (2) in this particular reading task, the students were asked to read orally -- thus they were compelled to attend to the graphics and match it with the appropriate sound

value. When reading silently, these average grade-four readers would probably infer the meaning and not identify the word.

To conclude, at the fourth-grade phase of learning to read, how grapho-phonics are processed appears to have little bearing on semantic processing. Furthermore, Goodman and Burke (1973) found, in their oral reading research with various groupings across elementary and high schools, that only the low grade-two readers showed "any definite inability or lack of confidence in using grapho-phonic information in reading" (p. 55).

Suggestions for Further Research

Suggestions for future research related to the conceptualization and assessment of reading as cognitive and linguistic processing will be discussed : first as broadly based issues related to the revision of the theory of reading processes ; and then as improvements to the assessment instrument, the ARP.

Revision of the Theory of Reading Processes

Although the foregoing factorial study procured some measure of construct validity for a process explanation of reading, the fit between the theory and the processes that actually occur during reading is still far from perfect. The results of the factorial study raise certain issues which must be resolved if the theory is to be further refined so as to narrow this gap.

(1) Rumelhart (1977) conceives of the reading processes as structured according to the hierarchical nature of text units, e.g.,

letters, words, etc ; hence, his processing model is hierarchical. In contrast, in the present theory unit of text is not considered, certain of the eight processes are measured across units of text, e.g., word synthesis, sentence-level synthesis and passage-level synthesis. But some aspects of the eight factor solution point to a possible breakdown of processing by unit of text. But it must be remembered that how the processes are measured at present may impose this hierarchical structure on them. However, if further research points to a processing-by-level explanation, some or all of the eight processes could be divided into distinct sub-processes. This issue has arisen as a result of comparing the present results to Rumelhart's (1977) model ; hence, other avenues of research will certainly be opened by comparing the present theory with alternate explanations of the reader-text interaction.

(2) The relationship between reading processes and the products or the dependent variables should be explored more closely. Although it has been demonstrated that certain processing configurations occur during reading, yet as has been shown for grapho-phonic processing, the relationship between the co-occurrance of processing and measures of reading comprehension is at best tenuous. More specifically, only one process, monitoring, loaded on the same factor as the conventional scholastic measure of reading output. The question is then how productive are the various processes for successfully completing different reading tasks ? Perhaps the root of the problem lies in differences between how the reading processes and products are measured.

(3) This theoretical speculation is closely tied to the

previous point. Processing information is collected by the ARP in two separate ways, namely, miscue and recall analysis. Miscue data is only available when the child makes an error during oral reading, hence it is only possible to infer certain processes such as attention to graphics or sentence-level synthesis when the child engages in some measure of "inexact" processing. Therefore the question that must be investigated is : What is the relationship between miscue processing and "exact" processing as it is measured by recalls and achievement tests ?

(4) As a result of carrying out the present research with a sample of average grade-four readers from Edmonton, a certain increment of validity has been established for the processing theory. There is, however, an urgent need to carry out similar research not only with various age and achievement levels but also with chidren from diverse ethnic and cultural backgrounds. Future work should build a wider nomological network by investigating the influence of such factors as the assessment situation, the examiner, and the directions on processing.

Revision of the ARP

Coding and Analysis. Because of the intimate link between a theory and how it is measured, some of the foregoing theoretical issues embody changes to the ARP, here however certain more specific changes will be introduced.

From the results of the factor analysis it appears that certain processes are more adequately assessed than others. Some measures may in fact be redundant, so could be eliminated, e.g., the indicators of sound-symbol association. Additional measures should

be sought for some processes, e.g., monitoring is only represented by one indicator -- corrected miscues. Finally, at present some measures are not distinct, i.e., the operational definition of some processes needs to be specified more exactly. For example, word accessing -- that is, immediate association of words -- should be divided into accessing of real and pseudo-words. In a similar manner, sequencing should be divided into the sequencing of words, which is related more to semantic/syntactic prediction, and the sequencing of graphemes within words, which is a dimension of either visual perception or orthographic prediction.

Passages. In the section devoted to item revision in Chapter 5, certain aspects of the stories such as conceptual load and vocabulary were modified as a result of field testing. The factorial validity study then represented the first tryout of the revised version of two new forms (C and D) ; at this stage certain changes to make the four forms parallel are inevitable. But in addition, the stories need to be rewritten so as to mitigate the stilted style resulting from writing the stories within the straightjacket of rigid controls. Perhaps some research could be undertaken to see the effect of presenting the stories in different prose styles.

Implications for Education

The presentation of implications resulting from the present research for educational practice will be divided into two sections : general implications, and more specific implications for the teaching of reading.

General Implications for Education

(1) Attempts should be made to establish the construct validity of all theories used in education. A knowledge of the logic of construct validity would help to mitigate the influence of fads and dogma that from time to time influence educational practice. This awareness would encourage teachers and school administrators to produce a theoretical rationale for current programs and teaching methods. Furthermore, new instructional methods would not be adopted without some evidence of their suitability for the particular population and age group in question. Also, authors and publishers of new teaching kits, reading series, workbooks, tests, etc., would then be obliged to establish a theoretical base and research evidence before the new products are distributed : a knowledgeable administrative and teaching population results in discriminating consumers.

(2) In all areas of educational measurement, tests or assessment instruments should be constructed by people with a knowledge of the basic discipline, e.g., mathematics or reading, and psychometrics ; then, and only then, will construct validity be established for educational measures. Furthermore, without a minimum of construct validity it is impossible to interpret the test scores in terms of cognitive processing ; and without a process explanation of test performance, it is difficult to translate the results into effective teaching strategies.

At present, many reading tests have been devised by people without a background in reading. In future this practice should definitely be avoided.

Implications for the Teaching of Reading

Since this study has established a certain measure of construct validity for a process explanation of reading, a process teaching orientation should prove successful in helping children to engage in effective reading processes. Besides, some recent publications (e.g., Goodman and Burke, 1980 ; Pearson and Johnson, 1978) have attempted to translate process theory into teaching strategies.

In fact, Goodman and Burke (1980) have devised specific lesson plans based on a psycholinguistic theory of reading. All these lessons, though focusing on various strategies such as predicting semantics and confirming, are aimed at making the reader aware of the purpose of reading. For instance, in the lessons on "predicting semantic cues" (pp. 43-59) the children are encouraged through the teacher's questioning to predict passage content based on the information they wish to derive from reading. Considering the importance of purpose in learning to read, these authors state :

> Too often students view reading as something that will help them to learn specific skills related to school instruction, but that is unrelated to their own need to comprehend the written language that surrounds them. They may not recognise that many of their daily contacts with written language outside the classroom (reading TV guides, comic books, the Bible, bubble-gum cards) are legitimate reading experiences that help them develop reading strategies (pp. 45-46).

Throughout the Goodman and Burke (1980) reading strategy lessons, every effort is exerted to help the children become active and involved readers, setting their own purposes, predicting meaning

and syntax, and constantly monitoring their progress.

In addition to psycholinguistics, Pearson and Johnson's guidelines for the teaching of reading comprehension are firmly based on the work of cognitive psychologists on discourse processing. Hence, these authors' suggestions for reading comprehension lessons are organized around the comprehension of various units of discourse : word, sentence, and passage. For instance, the children are encouraged to explore multiple word meaning and associations, deciphering a particular meaning by using the larger sentence and passage context. In some lessons on understanding longer discourse, the children are shown how to utilize the text structure to synthesize the printed message and abstract the main idea.

Despite the particular reading process lesson being taught, the children should be encouraged to use their experience and knowledge of language to become aware of blocks to comprehension ; but they should also be taught effective correction strategies such as re-reading or self-correction. All in all, the primary emphasis in reading instruction should be on comprehension rather than exact word identification ; in the present study, miscues which caused major disruptions in the communication of the author's message were related to an inability to synthesize the passage meaning rather than to the problem of utilizing grapho-phonic information. Thus, reading should be taught as active problem solving, where the reader is seen both as the recipient of meaningful communication and the constructor of meaning.

Concluding Statement

The primary purpose of the present work was to establish a measure of construct validity for the theory of assessment of reading process and the instrument designed to measure those processes, the ARP. To this end, theory and methods from two fields of knowledge -- psychometrics and reading -- were meshed. In synopsis, then, this research consisted of four main phases : a theory of reading process was outlined ; a model of test construction was developed ; an instrument capable of measuring reading process was designed and constructed ; and finally, a factorial validity study was undertaken.

Since the model of test construction which was developed invested construct validity with all the aspects of test construction and related evidence, three aspects of validational evidence were offered in the present work : item generation, item revision, and factorial validity. However, establishing construct validity is an ongoing enterprise, as the final evidence is never in ; this study was but the first step towards that goal.

BIBLIOGRAPHY

Adams, Jager M. Failures to comprehend and levels of processing in reading. Technical Report No. 37. Urbana, Illinois: University of Illinois, 1977.

Adams, Jager M. Models of word recognition. Technical Report No. 107. Urbana, Illinois: University of Illinois, 1978.

Allen, M.J. and Yen, W.M. Introduction to measurement theory. Monteray, California: Brooks/Cole Publishing Co., 1979.

Allington, R. Attention and application: The oft forgotten steps in teaching reading. Journal of Learning Disabilities, 1975, 8, 210-213.

American Psychological Association. Standards for educational and Psychological tests. Washington, D.C.: American Psychological Association, 1974.

Anastasi, A. Psychological testing, Second Edition. New York: Macmillan, 1968.

Anderson, R.C. How to construct achievement tests to assess comprehension. Review of Educational Research, 1972, 42, 145-170.

Anderson, R., Reynolds, R., Schallert, D., and Goetz, E. Frameworks for comprehending discourse. American Educational Research Journal, 1977, 14, 367-381.

Anderson, T.H., Wardrop, J.L., Hively, W., Muller, K.E., Anderson, R.I., Hastings, N.C., and Fredericksen, J. Development and trial of a model for developing domain referenced tests of reading comprehension. Technical Report No. 86. Urbana, Illinois, 1978.

Baker, E.L. Guidlines for preparing domains. In W. Hively (Ed.) Domain-referenced testing. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Baker, L. Comprehension monitoring: identifying and coping with text confusions. Journal of Reading Behavior, 1979, 11, 365-374.

Baker, L. and Anderson, R.C. Effects of inconsistent information on text processing: Evidence for comprehension monitoring. Reading Research Quarterly, 1982, 17, 281-294.

Bartlett, F.C. Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press, 1932.

Bateson, G. Steps to an ecology of the mind. New York: Ballantine Books, 1972.

Bechtoldt, H.P. Construct validity: a critique. In D.N. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Beebe, M.J. The effect of different types of substitution miscues on reading. Reading Research Quarterly, 1980, 15, 324-336.

Beebe, M.J. A model of the relationships between reading strategies and reading comprehension. Unpublished doctoral dissertation, University of Alberta, 1981.

Belth, M. The process of thinking. New York: David McKay, 1977.

Berk, R.A. (Ed.) Criterion-referenced measurement: the state of the art. Baltimore, MD.: Johns Hopkins University Press, 1980.

Betts, E.A. Foundations of reading instruction. New York: American Book Company, 1954.

Biemiller, A. The development of the use of graphic and contextual information as children learn to read. Reading Research Quarterly, 1970, 6, 75-96.

Biemiller, A. Relationships between oral reading rates for letters, words, and simple text in the development of reading achievement. Reading Research Quarterly, 1977, 13, 223-253.

Blalock, H.M. The measurement problem. In H.M. Blalock and A. Blalock (Eds.) Methodology in social research. New York: McGraw-Hill, 1968.

Bloomfield, L. Linguistics and reading. Elementary English, 1942, 19, 125-130.

Blum, J.M. Pseudo-science and mental ability. New York: Monthly Review Press, 1978.

Bower, G.H. Experiments on story understanding and recall. Quarterly of Experimental Psychology, 1976, 28, 511-534.

Brailsford, A. The relationship between cognitive strategy training and performance on tasks of reading with learning disabled children. Unpublished master's thesis. University of Alberta, 1981.

Brake, M.D. An investigation of oral and silent reading with low and high achievers. Unpublished master's thesis. University of Alberta, 1981.

Bransford, J.D. Human cognition: learning, understanding and remembering. Belmont, California: Wadsworth: 1979.

Brownowski, J. The ascent of man. London: British Broadcasting, 1973.

Brown, A.L. The development of memory: knowing about knowing, and knowing how to know. In H.W. Reese (ED.), Advances in child development (Vol. 10). New York: Academic Press, 1975.

Brown, A.L. Knowing when, where and how to remember: A problem of metacognition. In R. Glaser (Ed.), Advances in instructional psychology. Hillsdale, N.J.: Lawrence Erlbaum, 1978.

Brown, A.L. Metacognitive development and reading. In R.J. Spiro, B.C. Bruce and W.F. Brewer (Eds.), Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artifical intelligence and education. Hillsdale, N.J.: Lawrence Erlbaum, 1980.

Brown, A.L. Metacognition: The development of selective attention strategies for learning from texts. In M.L. Kamil (Ed.), Directions in reading: Research and instruction. Washington, D.C.: National Reading Conference, 1981.

Cambell, D.T. and Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. In D.N. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Carmines, E.G. and Zeller, R.A. Reliability and Validity Assessment. Sage University Paper series on quantative applications in the social sciences, 07-001. Beverly Hills: Sage Publications, 1979.

Carroll, J.B. On the theory - practice interface in the measurement of intellectual abilities. In P. Suppes (Ed.), Impact of research on education: Some case studies. Washington, D.C.: National Academy of Education, 1978.

Carroll, J.B. Defining language comprehension: some speculations. In J.B. Carroll and R.O. Freedle (Eds.), Language Comprehension and the acquisition of knowledge. Washington, D.C.: V.H. Winston, 1972.

Carroll, J.B., Davies, P. and Richman, B. Word frequency book. New York: American Heritage Publishing Co., Inc., 1971.

Cattell, R.B. Factor analysis: An introduction and manual for the psychologist and social scientist. New York: Harper and Brothers, 1952.

Cattell, R.B. Abilities: Their structure, growth and action. Boston: Houghton Miffin, 1971.

Chall, J.S. Learning to read: The great debate. New York: McGraw-Hill, 1967.

Chamberlain, T.C. Method of multiple working hypotheses. Science, 1965, 148, 755-759.

Chomsky, N. and Halle, M. The sound patterns of English. New York: Harper and Row, 1968.

Clarke, D. A comparison of the unaided recalls of able and less able readers. Unpublished master's thesis, University of Alberta, 1981.

Clark, H.H. and Clark, E.V. Psychology and Language. New York: Harcourt Brace Jovanovich, 1977.

Clay, M.M. A syntactic analysis of reading errors. Journal of Verbal Learning and Verbal Behavior, 1968, 7, 434-438.

Clay, M.M. Reading errors and self-correction behavior. British Journal of Educational Psychology, 1969, 39, 47-56.

Clay, M.M. Theoretical research and instructional change. In L.B. Resnick and P.A. Weaver (Eds.), Theory an practice of early reading, Volume 2, Hillsdale, N.J.: Lawrenece Erlbaum, 1979.

Cohen, G. The psychology of cognition. London: Academic Press, 1977.

Craik, F.I.M., and Lockhart, R.S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 671-684.

Cronbach, L.J. Five decades of public controversy over mental testing. American Psychologist, 1975, 30, 1-14.

Cronbach, L.J. Test validation. In R.L. Thorndike (Ed.), Educational measurement (2nd ed.). Washington, D.C.: American Council on Education, 1971.

Cronbach, L.J. and Meehl, P.E. (1955) Construct validity in psychological tests. In D.J. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Crothers, E.J. Inference and cohesion. Disourse Processes, 1978, 1, 51-71.

Das, J.P., Kirby, J.R. and Jarman, R.F. Simultaneous and successive cognitive processes. New York: Academic Press, 1979.

Davis, F.B. Measurement of improvement in reading skill courses. In E. Bliesmer (Ed.) Problems, programs and projects in college-adult reading. Eleventh yearbook of the National Reading Conference, Milwaukee: National Reading Conference, 1962.

Davis, F.B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

De Beaugrande, R. Design criteria for process models of reading. Reading Research Quarterly, 1981, 16, 261-315.

Dodd, D.H. and White, R.M. Cognition. Boston: Allyn and Bacon, 1980.

Drum, P.A. and Lantaff, R.E. Scoring categories for protocols. Paper presented at the Second Annual Language Conference, Boston, 1977.

Dubois, F.H. A history of psychological testing. Boston: Allyn and Bacon, 1970.

Eagan, R.L. An investigation into the relationship of the pausing phenomena in oral reading and reading comprehension. Unpublished doctoral dissertation, University of Alberta, 1973.

Eells, J. Intelligence and cultural difference, Chicago: The University of Chicago Press, 1951.

Enright, R.D. and Lapsley, D.K. Social role-taking: a review of the constructs, measures, and measurement properties. Review of Educational Research, 1980, 50, 647-674.

Estes, W.K. On the descriptive and explanatory functions of theories of memory. In L. Nilsson (Ed.), Perspective on memory research. Hillsdale, N.J.: Erlbaum, 1979.

Fagan, W.T. A comparison of the oral language of children ages nine, ten and eleven. A Research Report, Canada Council Grant, University of Alberta, 1978.

Fagan, W.T. Comprehension categories for protocol analysis. In W.T. Fagan, C. Cooper and J. Jensen (Eds.), Measures for Research and Evaluation in the English Language Arts, Volume 2. Urbana, Illinois: The National Council of Teachers of English, (in press).

Fagan, W.T. and Eagan, R.L. Word recognition strategies of children who made gains and children who did not make gains in a remedial reading program. Unpublished paper, University of Alberta, 1982.

Farr, R. Reading tests and teachers. Paper presented at the International Reading Association Conference, Boston, 1968.

Farr, R. Reading what can be measured? Newark, Delaware: International Reading Association, 1969.

Farr, R. The fallacies of testing. Paper presented to the Conference on Reading and the National Interest, Bloomington, Indiana, 1970.

Feuerstein, R. The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques. Baltimore: University Park Press, 1979.

Fitzgerald Whaley, J. Readers' expectations of story structure. Reading Research Quarterly, 1981, 17, 90-114.

Flavell, J.H. Cognitive monitoring. In W.P. Dickson (Ed.), Children's oral communication skills. New York: Academic Press, 1981.

Fodor, J.A., Bever, T.G., and Garrett, M.F. The psychology of language: An introduction to psycholinguistics and generative grammar. New York: McGraw-Hill, 1974.

Forester, D.C. The relationship of referential and logical text cues and the retention of information of average fourth grade readers. Unpublished master's thesis, University of Alberta, 1978.

Frederiksen, C.H. Semantic processing units in understanding text. In R.O. Freedle (Ed.), Discourse production and comprehension, Volume 1. Norwood, N.J.: Ablex Publishing, 1977.

Feifel, H., and Lorge, I. Qualitative differences in the vocabulary of children. Journal of Educational Psychology, 1950, 41, 1-8.

Feigl, J. Confirmability and confirmation. In P.P Wiener (Ed.), Readings in philosophy of science. New York: Scribner's, 1953.

Fleisher, L.S., Jenkins, J.R. and Pany, D. Effects on poor readers' comprehension of training in rapid decoding. Reading Research Quarterly, 1979, 15, 38-48.

Garner, R. amd Reis, R. Monitoring and resolving obstacles: an investigation of spontaneous text lookbacks among upper grade good and poor comprehenders. Reading Research Quarterly. 1981, 16, 569-582.

Gibson, E.J. Principles of perceptual learning and development. New York: Appleton-Century-Crofts, 1969.

Gibson, E.J. The ontogeny of reading. American Psychologist, 1970, 25, 136-143.

Gibson, E.J. Perceptual learning and the theory of word perception. Cognitive Psychology, 1971, 2, 351-368.

Gibson, E.J. Trends in perceptual development: impications for the reading process. In H. Singer and R. Ruddell (Eds.), Theoretical models and processes of reading. Newark, N.J.: International Reading Association, 1976.

Gibson, E.J., and Levin, H. The psychology of reading. Cambridge, Mass.: M.I.T. Press, 1975.

Gibson, E.J., Osser, H., Gibson, J.J. and Pick, A.D. A developmental study of the discrimination of letter-like forms. Journal of Comparitive and Physiological Psychology, 1962, 55, 897-906.

Gates, A.I. An experimental and statistical study of reading and reading tests. The Journal of Educational Psychology, 1921, 12, 303-314.

Gates, A.I. Bond, G.I. and Russell, D.H. Method of determining reading readiness. New York: Bureau of Publications, Teachers College, Columbia University, 1939.

Giboney, V. Communicative aspects of reading comprehension: A theoretical study of the role of intersubjectivity in written language communication. Unpublished doctoral dissertation, University of Alberta, 1979.

Golinkoff, R. Comprehension in good and poor readers. Reading Research Quarterly, 1976, 11, 625-659.

Goodman, K.S. _The psycholinguistic nature of the reading process_. Detroit: Wayne State University Press, 1968.

Goodman, K.S. Analysis of oral reading miscues: Applied psycholinguistics. _Reading Research Quarterly_, 1969, 5, 9-30.

Goodman, K.S. Strategies for increasing comprehension in reading. In H.M. Robinson (Ed.). _Improving reading in the intermediate years_. Chicago: Scott, Foresman, 1973.

Goodman, K.S. Influences of the visual peripheral field in reading. _Research in the teaching of English_, 1975, 9, 210-222.

Goodman, K.S. Reading a psycholinguistic guessing game. In H.S. Singer and R.B. Ruddell (Eds.), _Theoretical models and processes of reading_. Newark, Delaware: International Reading Association, 1976.

Goodman, K.S. After Miscue Analysis What? Paper presented at the meeting of NARSC/IRA, Edmonton, February, 1981.

Goodman, K.S. and Burke, C.L. _Theoretically based studies of patterns of miscues in oral reading performance_. Final Report, USOE Project No. 9-0375. Wayne State University, 1973.

Goodman, K.S. and Gollasch, F.V. Word omissions: deliberate and non-deliberate. _Reading Research Quarterly_, 1980, 16, 6-31.

Goodman, K.S. and Goodman, Y.M. Learning about psycholinguistic processes by analysing oral reading. _Harvard Educational Review_, 1977, 47, 317-333.

Goodman, Y.M. The Roots of Literacy. Paper presented at the Claremont Reading Conference, California, 1980.

Goodman, Y.M and Burke, C.L. _Reading miscue inventory_. New York: Macmillan, 1972.

Goodman, Y.M. and Burke, C.L. _Reading Strategies: Focus on comprehension_. New York: Holt, Rinehart and Winston, 1980.

Gough, P.B. One second of reading. In J.F. Kavanaugh and I.G. Mattingly (Eds.), _Language by eye and ear_. Cambridge, Mass.: M.I.T. Press, 1972.

Gould, S.J. _Ever since Darwin_. New York: Norton, 1977.

Gray, W.S. The major aspects of reading. In H.M. Robinson (Ed.), Sequential development of reading abilities. Chicago: Univerity of Chicago Press, 1960.

Guilford, J.P. The nature of human intelligence. New York: McGraw-Hill, 1967.

Hambleton, R.K., Algina, J. and Coulson, D.B. Criterion-referenced testing and measurement: a review of technical issues and developments. Review of Educational Research, 1978, 48, 1-47.

Haupt, E.J. and Goldsmith, J.S. Expanding the factorial structure of reading errors: frequency, severity, and higher-order components. Paper presented at the American Educational Research Meeting. New York, 1982.

Hempel, C.G. Philosophy of natural science. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

Henry, G.H. Teaching reading as concept development: Emphasis on affective thinking. Newark, Delaware: International Reading Association, 1974.

Hively, W. (Ed.) Domain-referenced testing. Englewood Cliffs, N.J.: Educational Technology Publications, 1974.

Hochberg, J. Components of literacy: speculations and explanatory research. In H. Levin and J.P. WIlliams (Eds.), Basic Studies in Reading. New York: Basic Books, 1970.

Hofstadter, R. Social Darwinism in American Thought. Philidelphia: University of Pennsylvania Press, 1944.

Holmes, J.A. The substrata-factor theory of reading: some experimental evidence. In H. Singer and R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1970.

Holmes, J.A. The substrata-factor theory of reading. Berkeley: California Book Company, 1953.

Hood, L. Qualitative analysis of oral reading errors: The interjudge reliability of scores. Reading Research Quarterly, 1976, 11, 557-598.

Horn, J.L On subjectivity in factor analysis. Educational and Psychological Measurement, 1967, 27, 811-820.

Huey, E.B. (1908). The psychology and pedigogy of reading. Cambridge: The M.I.T. Press, 1968.

Jackson, D.N. and Messick, S. (Ed.) Problems in human assessment. New York: McGraw-Hill, 1967.

Jenkinson, M.D. Basic elements in reading comprehension. In R. Staiger (Ed.), Reading a human right and a human problem. Newark, Delaware: International Reading Association, 1969.

Jenkinson, M.D. Selected processes and difficulties of reading comprehension. Unpublished doctoral dissertation. Department of Education, University of Chicago, 1957.

Johnson, M.K., Bransford, J.D. and Solomon, S. Memory for tacit implications of sentences. Journal of Experimental Psychology, 1973, 98, 203-205.

Johnson, M.S. and Kress, R.A. Informal reading inventories. Newark, Delaware: International Reading Association, 1965.

Juola, J.F., Schadler, M., Chabot, R., McCaughey, M. and Wait, J. What do children learn when they learn to read? In L.B. Resnick and P.A. Weaver (Eds.), Theory and practice of early reading, Volume 2. Hillsdale, N.J.: Lawrence Erlbaum, 1979.

Kaplan, R.I. The conduct of inquiry. New York: Basic Books, 1964.

Kavanaugh, M.B. Processing differences on three reading levels between no gain and gain groups in a remedial reading program. Unpublished master's project, University of Alberta, 1981.

Kemeny, J.G. A philosopher looks at science. Princeton, N.J.: Van Nostrand, 1959.

Kendall, J.R., Mason, J.H. and Hunter, W. Which comprehension? Artifacts in the measurement of reading comprehension. The Journal of Educational Research, 1980, 73, 233-6.

Killgallon, P.A. A study to determine relationships among certain pupils' adjustments in language situations. Unpublished doctoral dissertation, Pennsylvania State College, 1942.

Kim, J.O. and Mueller, C.W. Introduction to factor analysis. Sage University Paper series on quantiative applications in the social sciences, 07-013, Beverly Hills: Sage Publications, 1978.

Kingston, A.J. The measurement of reading comprehension. In R. Farr (Ed.), Measurement and evaluation of reading. New York: Harcourt, Brace and World, 1970.

Kintsch, W. and Van Dijk, T.A. Towards a model of text comprehension and production. _Psychological Review_, 1978, 85, 363-394.

Kintsch, W. amd Vipond, D. Reading comprehension and readability in educational practice and psychological theory. Paper presented at the Conference on Memory, University of Uppsala, 1977.

Kintsch, W. and Greene, E. The role of culture specific schemata in the comprehension and recall of stories. _Discourse Processes_, 1978, 1-13.

Kirsch, I.S. and Guthrie, J.T. Construct validity of functional reading tests. _Journal of Educational Measurement_, 1980, 17, 81-93.

Kuhn, T.S. (1962) _The structure of scientific revolutions_ (2nd edition). Chicago: University of Chicago Press, 1970.

Laberge, D. and Samuels, S.J. Toward a theory of automatic information processing in reading. _Cognitive Psychology_, 1974, 6, 293-323.

Labov, W. The logic of non-standard English. In J.E. Alatis (Ed.), _Report of the 21st annual round table meeting on linguistics and language studies_. Washington, D.C.: Georgetown University Press, 1970.

Labov, W., Cohen, P., Robins, C. and Lewis, J. _A study of the non-standard English of Negro and Puerto Rican speakers in New York City, volume 2_. Columbia University, Cooperative Research Project No. 3288.

Lehner, P.N. _Handbook of ethological methods_. New York: Garland, 1979.

Lennon, R.T. What can be measured. In R. Farr (Ed.), _Measurement and evaluation of reading_. New York: Harcourt, Brace and World, 1970.

Leu, D.J. Oral reading error analysis: A critical review of research and application. _Reading Research Quarterly_, 1982, 17, 420-437.

Levin, H. _The eye-voice span_. Cambridge, Mass.: M.I.T. Press, 1979.

Loevinger, J. Person and population as psychometric concepts. _Psychological Review_, 1965, 72, 143-155.

Loevinger, J. (1957) Objective tests as instruments of psychological theory. In D.J. Jackson and S. Messick (Eds.), Problems in human assessment. New York: McGraw-Hill, 1967.

Lorge, I. The fundamental nature of measurement. In D.N. Jackson and S. Messick, Problems in human assessment. New York: McGraw-Hill, 1967.

Lorge, I., Thorndike, R.L. and Hagan, E. The Lorge-Thorndike Intelligence Tests. Canadian Edition. Don Mills: Thomas Nelson and Sons, 1967.

Lupart, J.L. Attention and disabled readers: a top-down perspective. Unpublished doctoral dissertation, University of Alberta, 1981.

Luria, A.R. Higher cortical functions in man. New York: Basic Books, 1966.

MacGinitie, W.H., Kamons, J., Kowalski, R.L. MacGinitie, R.K. and MacKay, T. Gates-MacGinitie Reading Tests, Canadian Edition Teacher's Manual. Canada: Thomas Nelson and Sons, 1978.

Macleod, R.W. An explanatory study of inference and cognitive synthesis in reading comprehension with selected grade 4 readers. Unpublished doctoral dissertation, University of Alberta, 1978.

Maguire, T.O. Measurement and evaluation. In J.H.M Andrews and W.T Rogers (Eds.), Canadian research in education: a state of the art review. Report prepared for the Social Sciences and Humanities Research Council of Canada, 1981.

Malicky, G.M. A reading processes approach to diagnosis and remediation. Unpublished paper, University of Alberta, 1982.

Mandler, Jean M. "A Code in the Node: The use of story schema in retrieval. Discourse Processes, 1978, 1, 14-35.

Mandler, J.M. and Johnson, N.S. Rememberance of things passed: story structure and recall. Cognitive Psychology, 1977, 9, 111-151.

Massaro, D.W. Understanding language. New York: Academic Press, 1975.

Mattingly, I.G. Reading, linguistic awareness and language acquisition. Paper presented at the International Research Seminar on Linguistic Awareness and Learning to Read, Victoria, B.C., 1979.

McCullough, C.M. Responses of elementary school children to common types of reading comprehension questions. Journal of Educational Research, 1957, 51, 65-70.

McCracken, R.A. The development and validation of the standard reading inventory for the individual appraisal of reading performance in grades one through six. Proceedings of the International Reading Association, IX, Newark, Delaware: International Reading Association, 1964.

McCracken, R.A. The standard reading inventory. Klamath Falls, Oregon: Klamath Printing Company, 1966.

McDonald, A.S. Measuring reading performance. In R. Farr (Ed.), Measurement and evaluation of reading. NewYork: Harcourt, Brace and World, 1970.

Messick, S.A. The standard problem: Meaning and values in measurement and evaluation. American Psychologist, 1975, 30, 205-16.

Messick, S. Test validity and the ethics of assessment. American Psychologist, 1980, 35, 1, 012-1, 027.

Messick, S. Evidence and ethics in the evaluation of tests. Educational Researcher, 1981, 10, 9-20.

Meyer, B. The organization of prose and its effect on recall. Report No. 1 Reading and Learning Series, Ithica, New York: Cornell University Department of Education, 1974.

Minsky, M. A framework for the representation of knowledge. In P. Winston (Ed.), The psychology of computer vision. New York: McGraw-Hill, 1975.

Montagu, A. (Ed.) Race and IQ. New York: Oxford University Press, 1975.

Moore, L., Van Arsdale, P.W., Glittenberg, and Aldrich, R.A. The Biocultural basis of health. St. Louis: Mosby, 1980.

Nicholson, T. The relative effects of different error types on understanding of connected discourse, Unpublished doctoral dissertation, University of Minnesota, 1977.

Nitko, A.J. Distinguishing the many varieties of criterion-referenced tests. Review of Educational Research, 1980, 50, 461-485.

Nunnally, J.C. Psychometric theory. New York: McGraw-Hill, 1978.

Norman, D.A. and Bobrow, D.G. On the role of active memory processes in perception and cognition. In C.N. Cofer (Ed.), The structure of human memory. San Francisco: Frerman, 1976.

O'Donnell, R.C., Griffin, W.J. and Norris, R.C. Syntax of kindergarten and elementary school children: A transformational analysis. Champaign, Illinois: National Council of Teachers of English, 1967.

Olshavsky, J.E. Reading as problem solving: an investigation of strategies. Reading Research Quarterly, 1976, 12, 654-674.

Ortony, A. and Anderson, R. On putting apples into bottles: a problem of polysemy. Cognitive Psychology, 1975, 7, 167-180.

Pace, P. Interpreting qualitative oral reading error scores: Error patterns of seventh grade disabled and fourth grade normal readers. Paper presented at the 22nd Annual Convention of the International Reading Association, Miami Beach, 1977.

Peak, H. Problems of objective observation. In L. Festinger and D. Katz (Eds.), Research methods in the behavioral sciences. New York: Dryden Press, 1953.

Pearson, P.D. and Johnson, D.D. Teaching reading comprehension. Toronto: Holt, Rinehart and Winston, 1978.

Perfetti, C.A. Language comprehension and fast decoding: some psycholinguistic prerequisites for skilled reading comprehension. Paper presented to the Development of Reading Comprehension Seminar of the International Reading Assoiciation, Newark, Delaware, 1975.

Perfetti, C.A. and Hogoboam, T. The relationship between single word decoding and reading comprehension skill. Journal of Educational Psychology, 1975, 67, 461-469.

Phillips-Riggs, L. The relationship between reading proficiency, background knowledge, and inferencing strategies. Unpublished doctoral dissertation, University of Alberta, 1981.

Poffenberger, A.T. (Ed.) James McKeen Cattell man of science, Volume II, Addresses and Formal Papers. Lancaster, Penn.: The Science Press, 1947.

Popham, W.J. The development of criterion-referenced tests: technical considerations. Los Angeles: Instructional Objectives Exchange, 1976.

Popham, W.J. Criterion-referenced measurement. Englewood Cliffs, N.J.: Prentice Hall, 1978.

Popham, W.J. Domain specification strategies. In R.A. Berk (Ed.), Criterion-referenced measurement: the state of the art. Baltimore, Md.: John Hopkins University Press, 1980.

Popper, K.R. The logic of scientific discovery, London: Hutchinson, 1959.

Powell, W.R. Reappraising the Criteria for Interpreting Informal Reading Inventories. Proceedings of the International Reading Association, XIII, Part IV, Newark, Delaware: International Reading Association, 1969.

Rawson, H. Cognition and reading: an approach to instruction. In G.C. Mackennan and T.G. Waller (Eds.), Advances in Theory and practice, Volume 1. New York: Academic Press, 1979.

Resnick, L.B. and Weaver, P.A. (Ed.) Theory and practice of early reading, Volume 2. Hillsdale, N.J.: Lawrence Earlbaum, 1979.

Restak, R.M. The brain: the last frontier. New York: Warner Books, 1979.

Rubin, A.D. A Theoretical Taxonomy of the Differences Between Oral and Written Language. Technical Report No. 35, Center for the Study of Reading. Urbana, Illinois: University of Illinois, 1978.

Ruddell, R.B. Psycholinguistic implications for a system of communication model. In K.S. Goodman and J.T. Fleming (Eds.), Psycholinguistics and the teaching of reading. Newark, Delaware: International Reading Association, 1969.

Rumelhart, D.E. Notes on a schema for stories. In D.G. Bobrow and A. Collin (Eds.), Representation and understanding: Studies in cognitive science. New York: Academic Press, 1975.

Rumelhart, D.E. Toward an interactive model of reading. In S. Dornic (Ed.), Attention and performance, VI. Hillsdale, N.J.: Lawrence Earlbaum, 1977.

Samuels, S.J. Automatic decoding and reading comprehension. Language Arts, 53, 1976.

Shank, R.C. and Abelson, R.P. Scripts, plans, goals and understanding. Hillsdale, N.J.: Earlbaum, 1977.

Shoben, E.J. Theories of semantic memory: Approaches to knowledge and sentence comprehension. In R.J. Spiro, B.C. Bruce and W.F. Brewer (Eds.), Theoretical issues in reading comprehension. Hillsdale, N.J. Lawrence Earlbaum, 1980.

Simons, H.D. Linguistic Skills and Reading Comprehension. Unpublished doctoral dissertation, Harvard University, Cambridge, mass., 1970.

Singer, H. A developmental model of speed of reading in grades three through six. In H. Singer and R.B. Ruddell (Eds.), Theoretical models and processes of reading. Newark, Delaware: International Reading Association, 1970.

Smith, F. The role of prediction in reading. Elementary English, 1975, 52, 305-311.

Smith, F. Understanding reading. Toronto: Holt, Rinehart and Winston, 1978.

Smith, F. and Holmes, D.L. The independence of letter, word and meaning identification in reading. Reading Research Quarterly, 1971, 6, 394-415.

Smith, E.E. and Kleinman, G.M. Theoretical issues and instructional hints. In L.B. Resnick and P.A. Weaver (Eds.), Theory and Practice of Early Reading, Volume 2. Hillsdale, N.J.: Lawrence Earlbaum, 1979.

Smith, N.B. Graded selctions for informal reading diagnosis for grades 1 through 3. New York: University Press, 1959.

Smith, N.B. (1934) American reading instruction. Newark: International Reading Association, 1965.

Spiro, R. Remembering information from text: The "state of schema" approach. In R. Anderson, R. Spiro, and W. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, N.J.: Earlbaum, 1977.

Steffensen, M.S., Boag-Dev, C. and Anderson, R.C. A cross-cultural perspective on reading comprehension. Reading Research Quarterly, 1977, 15, 10-29.

Stein, N.T. How children understand stories: A developmental analysis. Centre for the Study of Reading. Technical Report No. 69. Urbana, Illinois: University of Illinois, 1978.

Stein, N.L. and Glen, C.G. An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), Multidisciplinary approaches to discourse comprehension. Hillsdale, N.J.: Alex, Inc., 1977.

Swenson, L.C. Theories of Learning. Belmont, California: Wadsworth, 1980.

Technical recommendations for psychological tests and diagnostic techniques. Psychological Bulletin Supplement, 1954, 51, 1-38.

Tierney, R.J., Bridge, C. and Cera, M.J. The discourse processing operations of children. Reading Research Quarterly, 1978, 14, 537-569.

Thorndike, E.L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 9, 323-332.

Thorndike, R.L. (Ed.) Educational measurement (Second Edition). Washington, D.C.: American Council on Education, 1971.

Thurstone, L.L. Multiple-factor analysis. Chicago, Illinois: University of Chicago Press, 1947.

Torgerson, W.S. Methods of scaling. New York: John Wiley and sons, 1958.

Traxler, A.E. A study of the Van Wagenen-Dvorak diagnostic examination of silent reading abilities. Eduational Records Bulletin No. 31. New York: Educational Records Bureau, 1941, 33-41.

Tuinman, J.J. Passage-Related Reading Comprehension Questions. Paper presented at the meeting of the American Eduacational Research Association, New York, 1971.

Turner, A. and Green, E. The construction and use of a propositional text-base. Technical Report No. 63. Institute for the Study of Intellectual Behavior, University of Colorado, 1977.

Venezky, R.L. and Massaro, D.W. The role of orthographic regularity in word recognition. In L.B. Resnick and P.A. Weaver, Theory and practice of early reading, Volume 1. Hillsdale, N.J.: Lawrence Erlbaum, 1979.

Walmsley, S.A. The criterion-referenced measurement of an early reading behavior. Unpublished doctoral dissertation, Harvard University, Cambridge, Mass., 1975.

Walmsley, S.A. The criterion-referenced measurement of an early reading behavior. Reading Research Quarterly, 1978, 14, 574-604.

Walter, R.B. History and Development of the Informal Reading Inventory. Unpublished study prepared at Kean College of New Jersey, 1974.

Ward, W.C., Frederiksen, N. and Carlson, S.B. Construct validity of free response and machine scorable forms of a test. Journal of Educational Measurement, 1980, 17, 11-29.

Wardrop, J.L., Anderson, T.H., Hively, W., Anderson, R.I. Hastings, N.C. and Muller, R.E. A framework for analysing reading test characteristics. Technical Report No. 109. Urbana, Illinois: University of Illinois, 1978.

Weber, R. A linguistic analysis of first-grade reading errors. Reading Research Quarterly, 1970, 5, 427-451.

Williams, B. and Clay, M.M The reading behavior of children in standard one. Education, 1973, 22, 13-17.

APPENDIX A

LEVEL 4 Form B

Budgies

Marty, Ann, and their little brother Rob live in an apartment. They would love to have a cat but their apartment building permits no pets except caged birds. Their uncle gave them one blue and two yellow budgies. Now they have only two because of little Rob's mischief. One sunny day last spring, Marty decided it would be nice to take their budgies outside to enjoy the fine weather. So he carried the cage very carefully and Ann led Rob by the hand. They found a quiet spot in the park and placed the cage on a park bench. They talked to their chirping budgie friends for a long time. Then Ann said, "Let's find twigs to decorate the cage." Marty and Ann searched under the big tree for dry twigs. At first, Rob sat beside the cage watching his brother and sister. Then he tried to get one of the budgies to play with. He got the cage door open, and in a flash a budgie escaped. When Rob began to sob loudly Ann and Marty came running, but by then the budgie was nowhere in sight. They did not scold Rob but simply carried the remaining two budgies safely home.

LEVEL 6 Form A

Gymnastics

I just met my three friends, Ann, John and Tom, coming out of the gym. They were excited and were talking continuously. They had spent the afternoon watching the junior gymnastics championships. Ann was particularly impressed with the floor gymnastics. It was very much like dancing, so gracious and light and accompanied by music. They were very interested in the trampoline and spent over an hour watching the boys and girls bounding several metres in the air, spinning, doing sommersaults, falling now on their feet, now on their stomachs, now on their backs. Gymnastics looked like a lot of fun although one young champion had a slight accident. A blond boy was doing a backward sommersault and he landed awkwardly on the edge of the trampoline. In an instant the boy was lying motionless on the floor. Ann, John and Tom sprang to their feet, immediately putting their first aid training into practice. Ann ran to the store room to find some blankets to keep him warm. John prevented the spectators from crowding around the boy. He kept them at a distance by telling them how important it was for the injured boy to have oxygen. Tom had raced to phone an ambulance before anyone else had time to think. In the meantime, the boy regained consciousness and tried to stand up. Although he seemed in good health, he was taken to the hospital for a check up. The officials thanked the three friends for their speedy assistance. All in all, it had been an exciting afternoon packed with activity. Then I overheard my friends say they wanted to join a club and see for themselves what gymnastics is really like.

APPENDIX B

OBLIQUE PROMAX ROTATION : EIGHT-FACTOR SOLUTION (1)

	h^2	1	2	3	4	5	6	7	8
Gates-Macginitie	0.562	0.707	-0.133	0.061	0.104	-0.177	0.010	0.289	-0.054
IQ Score	0.347	0.474	-0.146	0.049	0.092	-0.135	0.027	0.331	-0.079
No Meaning Change	0.665	0.188	-0.093	-0.162	0.708	-0.076	-0.011	0.035	-0.056
Minor Meaning Change	0.418	-0.053	0.188	-0.002	-0.503	0.084	0.152	-0.193	-0.448
Major Meaning Change	0.668	-0.188	-0.095	0.248	-0.344	0.032	-0.167	0.153	0.576
Corrected Miscues	0.165	0.465	-0.064	-0.033	-0.042	-0.051	0.165	0.116	0.032
Total Graphemes	0.650	-0.047	0.756	0.248	-0.078	0.111	-0.049	-0.070	0.071
Initial Graphemes	0.552	-0.180	0.706	0.177	-0.158	0.064	-0.070	-0.036	-0.004
Medial Graphemes	0.600	-0.159	0.652	0.218	-0.290	-0.057	-0.057	0.075	-0.003
Final Graphemes	0.538	0.065	0.539	-0.207	-0.395	0.067	0.012	-0.130	-0.239
Total Phonemes	0.592	-0.040	0.767	-0.126	0.005	0.027	-0.013	0.119	0.170
Initial Phonemes	0.532	-0.104	0.734	0.181	0.067	0.039	-0.102	0.167	0.130
Medial Phonemes	0.394	-0.210	0.514	0.308	0.341	0.101	-0.111	0.081	0.184
Final Phonemes	0.454	-0.132	0.508	-0.329	-0.240	0.028	0.076	-0.025	-0.169
Mis-Sequencing	0.570	-0.350	-0.285	-0.598	-0.955	0.052	-0.137	-0.074	-0.058
Correct Sequencing	0.678	0.361	0.178	0.689	0.206	-0.098	0.166	-0.033	0.085
Real Words	0.790	-0.315	0.085	0.330	0.852	-0.046	0.025	-0.032	0.038
Pseudo-Words	0.777	0.319	-0.028	-0.267	-0.855	0.028	-0.028	0.021	-0.030
Words Accessed	0.549	-0.185	0.015	0.663	-0.169	-0.001	-0.121	-0.055	-0.254
Partial Words Accessed	0.492	0.460	0.032	-0.432	-0.338	0.081	0.094	0.126	0.285
Phrases Accessed	0.550	-0.090	-0.149	-0.583	0.396	-0.023	0.042	0.005	0.020
Omissions	0.283	0.135	-0.098	0.212	0.290	0.113	0.012	0.400	0.205
Insertions	0.532	0.237	-0.141	-0.564	0.352	-0.085	0.066	-0.016	0.059
Uncorrected Miscues	0.761	-0.867	-0.152	0.117	0.010	0.044	-0.212	-0.132	0.051
Syntactic Acceptability at Sentence Level	0.362	0.062	0.272	-0.233	0.142	-0.105	-0.108	-0.344	-0.272

OBLIQUE PROMAX ROTATION : EIGHT-FACTOR SOLUTION (2)

	h^2	1	2	3	4	5	6	7	8
Partial Sentence Syntactic Accept.	0.328	0.157	0.172	-0.046	-0.086	-0.077	-0.034	0.562	0.065
Semantic Accept. at Sentence Level	0.818	0.053	0.048	-0.247	0.815	0.041	-0.095	-0.092	-0.072
Partial Sentence Acc.	0.805	-0.006	-0.104	-0.025	0.876	0.107	0.088	0.137	0.198
Passage Semantic Acc.	0.585	0.304	0.066	-0.248	0.272	-0.063	0.031	-0.244	0.461
Sentence Prediction	0.829	0.024	0.085	-0.294	0.793	0.079	-0.072	-0.174	-0.134
Partial Sent. Predict.	0.830	-0.023	-0.118	-0.069	0.883	0.057	0.088	0.117	0.098
Oral Exact Recall	0.596	-0.385	0.172	0.095	0.091	0.784	-0.173	-0.089	-0.062
Oral Specific Recall	0.525	-0.131	0.054	-0.064	-0.044	-0.075	-0.504	0.423	-0.221
Oral Entailed Recall	0.533	0.282	-0.220	-0.019	-0.053	-0.045	0.753	0.148	-0.101
Oral Experiential Rec.	0.450	0.017	0.009	0.041	-0.137	-0.031	-0.057	-0.668	-0.139
Oral Elaborating Rec.	0.602	0.010	0.191	-0.113	0.211	0.050	0.243	0.013	0.709
Oral Error 1 Recall	0.352	-0.176	0.116	0.023	0.093	-0.446	-0.227	0.297	0.269
Oral Error 2 Recall	0.735	0.049	0.010	-0.142	-0.037	-0.037	-0.167	0.011	0.878
Oral Exact Input	0.579	-0.393	0.161	-0.014	0.107	0.718	-0.313	-0.186	-0.073
Oral Specific Input	0.637	-0.290	0.046	-0.118	0.074	0.015	-0.772	0.426	-0.265
Oral Entailed Input	0.153	-0.101	0.078	-0.193	-0.056	0.005	0.512	-0.193	-0.318
Oral Experiential Inp.	0.452	-0.317	0.079	-0.044	0.076	-0.118	-0.368	-0.592	-0.163
Oral Elaborating Inp.	0.650	0.013	0.225	-0.127	0.229	0.061	0.175	-0.008	0.757
Oral Error 1 Input	0.502	-0.189	0.082	0.028	0.062	-0.412	-0.286	0.425	0.372
Oral Error 2 Input	0.756	0.066	0.013	-0.156	0.014	-0.060	-0.158	-0.062	0.857
Silent Exact Recall	0.477	-0.037	0.069	-0.052	-0.030	0.674	0.112	0.085	0.135
Silent Specific Rec.	0.536	0.290	-0.074	0.253	0.037	0.089	-0.650	0.462	0.010
Silent Entailed Rec.	0.299	0.307	-0.048	-0.095	0.030	-0.098	0.563	0.030	-0.093

OBLIQUE PROMAX ROTATION : EIGHT-FACTOR SOLUTION (3)

	h^2	1	2	3	4	5	6	7	8
Silent Experiential Recall	0.519	0.070	-0.081	-0.050	0.050	-0.369	-0.207	-0.470	0.123
Silent Elaborating Rec.	0.493	-0.211	0.078	-0.099	-0.094	0.128	0.010	-0.700	-0.065
Silent Error 1 Recall	0.426	-0.463	0.191	-0.070	-0.025	-0.098	0.021	0.330	-0.162
Silent Error 2 Recall	0.666	-0.653	-0.208	-0.268	0.104	-0.044	0.235	0.080	-0.086
Silent Exact Input	0.513	-0.072	0.097	-0.113	0.014	0.632	-0.320	0.239	0.221
Silent Specific Input	0.654	0.183	0.019	0.128	0.050	0.259	-0.553	0.460	0.117
Silent Entailed Input	0.129	0.337	-0.235	-0.064	0.033	-0.201	0.093	0.077	0.031
Silent Experiential Inp.	0.626	0.147	-0.143	0.052	-0.054	-0.246	-0.348	-0.538	0.245
Silent Elaborating Inp.	0.624	-0.303	0.011	-0.040	-0.128	0.075	-0.106	-0.745	0.011
Silent Error 1 Input	0.378	-0.464	0.230	-0.099	0.010	-0.050	-0.199	0.335	-0.103
Silent Error 2 Input	0.521	-0.648	0.087	-0.261	0.027	-0.097	0.024	0.072	-0.124
Passage Level Read	0.466	0.650	0.057	0.182	0.013	-0.202	0.040	0.064	-0.053
PERCENTAGE COMMON VARIANCE		15.501	12.142	11.039	19.429	8.177	9.675	11.982	12.081
PERCENTAGE TOTAL VARIANCE		8.447	6.617	6.015	10.587	4.456	5.272	6.529	6.583